Journal of Clinical Child & Adolescent Psychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hcap20

Meta-Analysis of Early Intensive Behavioral Intervention for Children With Autism

Sigmund Eldevik [a b c] , Richard P. Hastings [a] , J. Carl Hughes [a] , Erik Jahr [d] , Svein Eikeseth [b] & Scott Cross [e]

[a] School of Psychology, Bangor University,

[b] Faculty of Behavioral Science, Akershus University College,

[c] Highfield Centre,

[d] Akershus University Hospital,

[e] Lovaas Institute for Early Intervention,

Published online: 13 May 2009.

To cite this article: Sigmund Eldevik , Richard P. Hastings , J. Carl Hughes , Erik Jahr , Svein Eikeseth & Scott Cross (2009) Meta-Analysis of Early Intensive Behavioral Intervention for Children With Autism, Journal of Clinical Child & Adolescent Psychology, 38:3, 439-450, DOI: 10.1080/15374410902851739

To link to this article: http://dx.doi.org/10.1080/15374410902851739

Journal of Clinical Child & Adolescent Psychology, 38(3), 439–450, 2009

ISSN: 1537-4416 print/1537-4424 online
DOI: 10.1080/15374410902851739

Routledge
Taylor & Francis Group

Meta-Analysis of Early Intensive Behavioral Intervention for Children With Autism

Sigmund Eldevik
School of Psychology, Bangor University; Faculty of Behavioral Science, Akershus University College; and Highfield Centre

Richard P. Hastings and J. Carl Hughes
School of Psychology, Bangor University

Erik Jahr
Akershus University Hospital

Svein Eikeseth
Faculty of Behavioral Science, Akershus University College

Scott Cross
Lovaas Institute for Early Intervention

A systematic literature search for studies reporting effects of Early Intensive Behavioral Intervention identified 34 studies, 9 of which were controlled designs having either a comparison or a control group. We completed a meta-analysis yielding a standardized mean difference effect size for two available outcome measures: change in full-scale intelligence and/or adaptive behavior composite. Effect sizes were computed using Hedges's *g*. The average effect size was 1.10 for change in full-scale intelligence (95% confidence interval = .87, 1.34) and .66 (95% confidence interval = .41, .90) for change in adaptive behavior composite. These effect sizes are generally considered to be large and moderate, respectively. Our results support the clinical implication that at present, and in the absence of other interventions with established efficacy, Early Intensive Behavioral Intervention should be an intervention of choice for children with autism.

There is a developing evidence base for the positive effects of comprehensive interventions for children with autism spectrum disorders (ASD). Two recent narrative reviews have focused on a range of comprehensive interventions for children with autism (Eikeseth, 2009; Rogers & Vismara, 2008). The conclusion from both of these reviews is that Early Intensive Behavioral Intervention (EIBI) is an effective intervention when compared against no intervention controls or eclectic/autism-specific special education interventions. When applying more formal criteria (Chambless et al., 1998; Chambless & Hollon, 1998; Chambless et al., 1996), Rogers and Vismara found that EIBI (or what they call the "Lovaas treatment approach") should be considered "well established" and that no other intervention presently qualifies for this status.

We thank Dr. Michael Borenstein for helpful advice in carrying out the present analysis. We also thank Alison Sharland and Iwona Ondire for coding papers. Furthermore we express our gratitude to the Sharland Foundation for a research grant supporting this work. This paper was submitted as part of the first author's Ph.D. thesis at Bangor University.

Correspondence should be addressed to Sigmund Eldevik, Akershus University College, P. O. Box 423, N-2001 Lillestrøm, Norway. E-mail: eldevik@online.no

EIBI programs (including the Lovaas treatment approach) have been described by Green, Brennan, and Fein (2002, p. 70, see also Eikeseth, 2009, for a similar definition) as having the following common elements: (a) intervention is individualized and comprehensive, addressing all skill domains; (b) many behavior analytic procedures are used to build new repertoires and reduce interfering behavior (e.g., differential reinforcement, prompting, discrete-trial instruction, incidental teaching, activity-embedded trials, task analysis, and others); (c) one or more individuals with advanced training in applied behavior analysis and experience with young children with autism directs the intervention; (d) normal developmental sequences guides the selection of intervention goals and short-term objectives; (e) parents serve as active co-therapists for their children; (f) intervention is delivered in one-to-one fashion initially, with gradual transitions to small-group and large-group formats when warranted; (g) intervention typically begins in the home and is carried over into other environments (e.g., community settings), with gradual, systematic transitions to preschool, kindergarten, and elementary school classrooms when children develop the skills required to learn in those settings; (h) programming is intensive, is year round, and includes 20 to 30 hr of structured sessions per week plus informal instruction and practice throughout most of the children's other waking hours; (i) in most cases, the duration of intervention is 2 or more years; and (j) most children start intervention in the preschool years, when they are 3 to 4 years of age.

In addition to narrative reviews, there have been two recent systematic reviews of outcome research on EIBI. The first review presented a systematic description of the research published to date and pointed to challenges for future research (Howlin, Magiati, & Charman, in press). Eleven studies were identified using the following inclusion criteria: The study had to have a control or comparison group with a minimum of 10 participants in each group, participant at intake had to be younger than 6 years of age, and intervention had to be provided for at least 12 hr a week for 12 months. Howlin et al. discussed a number of problems associated with drawing conclusions about the efficacy of EIBI. First, although not accurately reported in some of the studies, they estimated that the EIBI groups on average received significantly more hours of intervention than did control groups. Second, a variety of assessment instruments were used across children and studies that made it difficult to compare results across studies and may have led to results being spuriously positive. Third, in some studies it was unclear at what points in time the assessments were conducted, particularly at posttreatment when in some cases assessments were undertaken years after treatment had ended. Fourth, the studies reported test scores in different ways that including standard scores, age equivalents, and raw scores.

Howlin et al. (in press) concluded that in general the average effects of EIBI were favorable compared to controls but that the variability across individual children in the EIBI studies was substantial. Howlin et al. could not identify any reliable predictors of outcome. Intake IQ was found by some researchers to be related to better outcomes but others found no such relationship. Furthermore, age at intake was not found to be related to outcome in any of the studies. However, age range was limited with all children being younger than 7 years of age. Initial language ability was identified as a possible predictor only in some of the studies that explored this, and autism symptomatology was found to be related both with better and with worse outcomes, in two different studies. Given these problems they concluded that conducting a meta-analysis of the evidence was not appropriate.

The second recent systematic review conducted by Reichow and Wolery (2009) addressed similar questions to Howlin et al. (in press) and drew similar conclusions. However, unlike Howlin et al., they included a meta-analysis. The authors argued that a meta-analysis of EIBI is feasible but that it had to be limited to change in intelligence scores and that to have enough studies, they would have to include studies that were not controlled. Thus, the meta-analysis used standardized mean change effect sizes and not the more methodologically rigorous standardized mean difference effect size. The mean change effect size is computed without comparison or control group data and, as the authors point out, any conclusions are limited by threats to validity such as maturation. In addition, the standardized mean change effect size may inflate effect size estimates (Morris, 2000). Based on 12 studies, Reichow and Wolery reported a weighted mean effect-size for change in intellectual functioning following EIBI of .69.

A second aspect of the Reichow and Wolery (2009) analysis that may affect the validity of conclusions was that studies using a variety of outcome measures for intelligence were included. For example, studies which primarily relied on performance based nonverbal measures of intelligence such as the Merrill-Palmer (Stutsman, 1948) and the Leiter-R (Roid & Miller, 1997) were treated as equivalent to studies which reported full-scale IQ measures. Because the performance-based tests measure areas where children with autism often are relatively strong (e.g., visual-spatial tasks), scores tend to be higher than on full-scale IQ tests (Lord et al., 2006). Not separating these tests in an analysis on the effects of EIBI may affect the conclusions drawn, especially in cases where these different measures are used interchangeably pre- and post-treatment. A minor methodological problem with the Reichow and Wolery analysis was that they reported reliability estimates only on the coding of information from the selected studies (research methods, participants, and intervention characteristics); no such data were provided for the initial procedure for selecting studies to be included in the review.

The purpose of our study was to provide a replication and extension of the Reichow and Wolery (2009) meta-analysis, with a focus on methodological improvements. First, we selected studies with comparison/control groups only, while employing a more precise definition of EIBI (Green et al., 2002) and the control/comparison groups. This makes it possible to apply more methodologically rigorous mean difference effect size measures. Second, we required more uniformity in outcome measurement and included only full-scale measures of intelligence. Third, we were able to add a meta-analysis of changes in adaptive behavior. Fourth, because our literature search was conducted later in time and employed a somewhat different definition of EIBI, some additional recent published studies were included. Fifth, we included interrater reliability for our literature search and initial selection procedure for studies in the review. Sixth, we based our analysis on individual raw data gathered from authors rather than group average data reported in the original papers. This made it possible to prevent children from being represented more than once if they were included in more than one published outcome study and to ensure the selection of evaluation periods as similar as possible across studies.

METHODS

Search Strategy and Data Collection

We searched the PsycINFO, PubMed, and ERIC databases (up to March 2008) using a combination of the following terms: *behavior analytic, behavioral, early, intervention*, and *autism and/or PDD*. The first author read the headings and abstracts of all the papers from this initial search to decide whether the study warranted a more detailed coding. If it was possible that the study reported outcome data on the effects of behavioral treatment for children with ASD, the study was obtained for more detailed coding. In addition, the reference sections of obtained papers were browsed in an attempt to locate studies that might have been missed in the electronic search.

A coding scheme was used for coding all the selected studies (available from the first author). First, it was coded whether the children had received behavioral intervention that generally adhered to common elements described by Green et al. (2002, p. 70). Second, a series of true/false scores were given for the following: (a) the participants were on average between 2 and 7 years of age when intervention started; (b) the children were independently diagnosed with autism or PDD-NOS; (c) a full-scale measure of intelligence and/or a standardized measure of adaptive behavior such as the Vineland Adaptive Behavior Scales (VABS; Sparrow, Balla, & Cicchetti, 1984) was conducted at intake and after intervention (primarily administering a nonverbal intelligence measure such as the Leiter–R [Roid & Miller, 1997] or the Merrill–Palmer Scale of Mental Tests [Stutsman, 1948] led to the study's exclusion); (d) the duration of intervention was between 12 and 36 months; (e) the study was not a case study (or series of case studies); (f) the results had been published in a peer-reviewed journal; and (g) the study included either a control or comparison group. The studies were classified as either a comparison or a control study (or both). If it was specified that the children in the study had received intervention(s) other than EIBI of similar duration and intensity in terms of 1:1 hours, it was classed as a comparison study. Although it would probably be impossible to determine whether the children in the comparison groups had similar eclectic or specialist autism provision (even within a single study), classifying the studies in this way could still yield useful information. For example, it may facilitate the exploration of whether it is the number of 1:1 hours itself (i.e., "intensity") that makes a difference. Where no intervention (or a considerably less intensive one) was provided, or a poorly specified intervention was described, the study was classified as a control study.

The electronic and manual searches resulted in 2,150 potential hits. Through the first screening process, we selected 34 papers for detailed coding. One of the database searches, resulting in 607 hits, was chosen for a reliability check. The screening results from the first author were compared to that of a second screener (another author) using the same decision criteria. Agreement was high overall in terms of whether the paper should be subject to further analysis (Cohen's $\kappa = .85$). Disagreements occurred only because the second screener included fewer studies than the first screener. Thus, there were no instances of the second screener including a study for further analysis that was not already included by the first screener.

The 34 studies that remained after this initial screening were coded by the first author and two independent scorers (master's students in behavior analysis). Agreement was calculated between the first author and each of the independent scorers separately by dividing the total number of agreements by the total number of agreements plus disagreements and multiplying by 100. Initial agreement was high in both cases (91% and 94%, respectively) and the few disagreements that occurred were resolved after brief discussions. We excluded 25 of the 34 studies for one or more of the following reasons: (a) seven had inadequate intake and/or outcome data, such as primarily reporting performance IQ instead of full-scale IQ (Bibby, Eikeseth, Martin, Mudford, & Reeves, 2002; Drew et al., 2002; Fenske, Zalenski, Krantz, & McClannahan, 1985; Luiselli, Cannon, Ellis, & Sisson, 2000; Magiati, Charman, & Howlin, 2007; Sheinkopf & Siegel, 1998; Solomon, Necheles, Ferch, & Bruckman, 2007); (b) five had an intervention duration that was too short to meet inclusion criteria (Harris, Handleman, Gordon, Kristoff, & Fuentes, 1991; Ingersoll, Schreibman, & Stahmer, 2001; Reed, Osborne, & Corness, 2007a, 2007b; Stahmer & Ingersoll, 2004); (c) two reported data from case studies only (Butter, Mulick, & Metz, 2006; Green et al., 2002); (d) three reported data that were already included in other studies (Beglinger & Smith, 2005; Eikeseth, Smith, Jahr, & Eldevik, 2007; McEachin, Smith, & Lovaas, 1993); (e) upon closer inspection, one of the studies provided intervention that did not meet the definition of behavioral treatment (Gabriels, Hill, Pierce, Rogers, & Wehner, 2001); and (f) seven did not have a control or comparison group (Anderson, Avery, DiPietro, Edwards, & Christian, 1987; Ben-Itzchak & Zachor, 2007; Harris & Handleman, 2000; Hayward, Eikeseth, Gale, & Morgan, in press; Sallows & Graupner, 2005; Smith, Buch, & Gamby, 2000; Weiss, 1999).

Individual data needed to calculate effect sizes from the nine remaining studies were obtained by contacting the authors of each study. We asked them to provide the age, IQ, and adaptive behavior scores at intake and after 2 years in intervention (or as close as possible). Also, we asked if any of the children either in the EIBI or comparison/control groups was represented in other published studies. Thus, all computations in our study were conducted by recalculating pre- and postgroup means and standard deviation on outcome measures rather than data reported in the original papers or extrapolated from these reports. Individual data from the second control group ($n = 21$) in the Lovaas (1987) study were not available, and 4 children in the comparison group from one study (Eldevik, Eikeseth, Jahr, & Smith, 2006) were excluded because they were already in the comparison group of an earlier study also included in the analysis (Eikeseth, Smith, Jahr, & Eldevik, 2002). Figure 1 summarizes the study search and selection process.

The total number of children in the nine intervention studies was 297–153 in the EIBI groups, 105 in control groups and 39 in comparison groups. Table 1 summarizes the main characteristics of the children included in this analysis, including mean age at intake, IQ, and Vineland Adaptive Behavior Composite (ABC) scores at intake and posttreatment. Not all authors reported both IQ and ABC data, or were able to give the exact duration of intervention for each individual child. The average

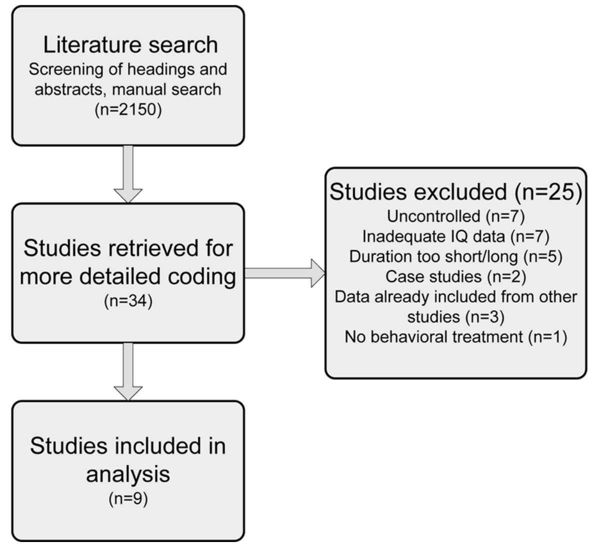

FIGURE 1 The search and selection procedure.

intensity in terms of weekly hours and duration is provided in Table 1. The research design and assignment procedures employed are briefly described along with any inclusion criteria described in the original paper. If a study reported outcome data at more than one point in time, we chose the point that was closest to 2-year duration of treatment. All of the aforementioned calculations were conducted in SPSS (version 16.0) using raw data provided from the authors. Hence, the pre- and postgroup means and standard deviations may differ from those reported in the original published papers.

Child Measures

Intellectual functioning. The Bayley Scales of Infant Development, either the first or second edition (Bayley, 1969, 1993) were used for the youngest children or the children that scored below the basal on intelligence tests standardized for their chronological age. The Bayley Scales of Infant Development is a measure of mental development for children up to 42 months. It will yield a mental developmental index, which is considered broadly equivalent to an IQ score. For the older and higher functioning children the most frequently used measures of intelligence were the Stanford–Binet Intelligence Scale, Fourth Edition (Thorndike, Hagen, & Sattler, 1986), the Wechsler Preschool and Primary Scale Intelligence–Revised (Wechsler, 1989), the Wechsler Intelligence Scale for Children–Revised (Wechsler, 1974), or the Wechsler Intelligence Scales for Children–Third Edition (Wechsler, 1993). If the child scored below the norms on a test, researchers had generally computed a ratio IQ score by dividing the obtained mental age with chronological age and multiplying by 100. All of the tests have been used extensively and validated for children

TABLE 1
Main Characteristics of the Studies Included in the Analysis of Outcome. Studies Included in Other Recent Reviews are Shown

			Pre		Post		Intensity				
Study	*Group*	*Age*	*IQ*	*ABC*	*IQ*	*ABC*	*hrs*	*mths*	*n*	*Design/Assignment/ Inclusion*	*Comments*
USA											
(Lovaas, 1987) 1, 2, 3, 4	EIBI	34.6 (8.9)	62.9 (13.7)		83.3 (28.6)		40	24–36	19	QCT/Staff availability and archives. Included if CA < 40 months if mute or CA < 46 months if echolalic and prorated mental age of >11 months at CA 30 months.	Five subjects deemed untestable at intake, 3 in experimental group and 2 in control group 1. Intelligence scores based on mental age scores from the Vineland Social Maturity Scale (Doll, 1953) were used in these cases.
	Control	40.9 (10.3)	57.1 (14.5)		50.1 (22.4)		<10	24	19		
Smith, Eikeseth, Klevstrand, & Lovaas, 1997) 1, 3, 4	EIBI	36.0 (6.9)	27.8 (4.9)	50.3 (9.1)	35.8 (14.3)	51.7 (17.9)	30	24	11	QCT/Archival data. Included if CA < 46 months and IQ < 35	Post measures conducted 3–4 years after treatment in some cases. VABS data only available for 6 of the 11 children in the ABA group. Control group received minimal treatment.
	Control	38.0 (5.4)	27.3 (5.4)		24.0 (8.2)		<10	24	10		
(Howard, Sparkman, Cohen, Green, & Stansislaw, 2005) 2, 3	EIBI	30.9 (5.2)	58.5 (18.2)	70.5 (11.9)	89.9 (20.9)	81.3 (11.1)	25–40	14	29	QCT/Parental preference and IEP teams. 16 Included if CA < 48 months.	EIBI: Multiple settings (home, school and community. 25–30 h per week under 3 years of age, 35–40 h per week over 3 years of age. AP (autism educational programming): Public classroom for children with autism. 1:1 or 1:2 staff: child ratio. 25–30 hrs per week of intervention, supervision by special education teacher. Intervention eclectic (PECS, SIT; TEACCH, DTT). 7 children received 1–2 session per week of speech therapy. GP (generic educational programming): Local community special education classrooms. Average of 15 hrs per week intervention, 1:6 staff: child ratio. 13 children received speech and language therapy for 1–2 times per week.

(*Continued*)

TABLE 1
Continued

Study	*Group*	*Age*	*Pre*		*Post*		*Intensity*		*n*	*Design/Assignment/ Inclusion*	*Comments*
			IQ	*ABC*	*IQ*	*ABC*	*hrs*	*mths*			
	Comparison (AP)	37.4 (5.7)	53.7 (13.5)	69.8 (10.5)	62.1 (19.6)	69.1 (12.9)	25	13	16		
	Control (GP)	34.6 (6.5)	59.9 (14.8)	71.6 (10.5)	68.8 (15.3)	68.3 (9.9)	15	15	16		
(Smith, Groen, & Wynn, 2000) 1, 2, 3, 4	EIBI	36.1 (6.0)	50.5 (11.2)	63.7 (9.6)	66.5 (24.1)	61.3 (28.7)	24.5	24	15	RCT/Matched-pair random. Included if CA < 42 months and ratio IQ between 35 and 75.	Clinic directed group: Number of hours for ABA group are for first year in treatment. Gradual reductions in year two. Treatment phased out after 18 months for children responding slowly. Average duration 33 months. Parent managed group: 5 hrs a week of parent training for first 3–9 months, parents asked to do 5 hours a week in between sessions: Total <10 hours per week of ABA +12.5 hours of special education classes per week. ABA treatment hours second year presumed to be gradually decreasing, school hours presumed to be the same. Follow-up testing at CA 7–8 years. Duration between testing on average 54 months. Autism and PDD-NOS lumped together in the present analysis.
	Comparison	35.7 (5.4)	50.7 (13.9)	65.2 (9.0)	50.5 (20.4)	59.9 (16.7)	<10	~24	13		
(Cohen, Amerine-Dickens, & Smith, 2006) 1, 2, 3, 4	EIBI	34.4 (5.4)	62.0 (16.4)	64.0 (8.4)	81.1 (21.8)	79.5 (13.4)	35–40	24	21	QCT/Parental preference. Included if CA < 48 months and ratio IQ > 35.	Community-non-university setting. Community services selected by family. In control group 1 child had a Early Start Autism Intervention Program 9 hrs a week, 2 children home-based development program 1–4 hrs a week, 17 SDC (special day class) eclectic, ratio 1:1 to 3:1, 3–5

											days a week for up to 5 hrs. Speech, behavioral and occupational therapies 0–5 hrs per week. 3 where mainstreamed for up to 45 minutes a day.
	Control	33.2 (3.7)	59.4 (14.7)	71.9 (11.5)	65.9 (16.5)	70.7 (13.3)	–	24	21		
Australia (Birnbrauer & Leach, 1993) 1, 2, 4	EIBI	38.1 (7.1)	45.3 (17.9)	47.5	57.6 (18.7)	41.0 (5.1)	19	22	9	QCT/Parent willingness and geographical. 5 Included if CA between 24 and 48 months.	Untestable subjects set to IQ of 30. Ratio scores computed for the rest of subjects. Scores post treatment are deviation IQ used where available. Ratio VABS scores calculated both pre and post.
	Control	33.2 (10.3)	45.0 (9.4)	51.5	43.2 (15.0)	42.5 (4.9)	–	22	5		
Norway (Eikeseth, Smith, Jahr, & Eldevik, 2002) 1, 2, 3, 4	EIBI	66.3 (11.3)	61.9 (11.3)	55.8 (9.0)	79.1 (18.1)	67.0 (16.3)	28	12	13	QCT/Staff availability. Included if CA between 48 and 84 months and IQ ≥ 50	Comparison received eclectic treatment of similar intensity, but with less supervsision.
	Comparison	64.8 (9.9)	65.2 (15.0)	60.0 (13.2)	68.9 (18.8)	60.2 (11.7)	29	14	12		
(Eldevik, Eikeseth, Jahr, & Smith, 2006) 1, 2, 3, 4	EIBI	53.1 (9.5)	41.0 (15.2)	52.5 (3.9)	49.2 (16.6)	52.4 (9.2)	13	20	13	QCT/Archival data. Included if CA < 72 months.	Comparison received eclectic treatment of similar intensity. Four subjects from comparison group taken out here, because included in Eikeseth et al. (2002).
	Comparison	45.1 (16.5)	42.8 (13.0)	50.1 (9.2)	38.5 (15.5)	44.6 (7.5)	12	23	11		
United Kingdom (Remington et al., 2007) 2, 3	EIBI	35.7 (4.0)	61.4 (16.7)	60.2 (5.8)	73.5 (27.3)	61.5 (15.4)	26	24	23	QCT/Parent preference. Included if CA between 30 and 42 months.	Control group received TAU, special school, mainstream or mix, but little or no 1:1, speech therapy, TEACCH etc.
	Control	38.4 (4.4)	62.3 (16.6)	57.0 (6.8)	60.1 (27.8)	54.6 (13.1)	16	24	11		

QCT = Quasi-experimental controlled clinical trial, RCT = Randomized controlled clinical trial.
1. Included in Reichow & Wolery (2009); 2. Included in Eikeseth (2009); 3. Included in Magiati, Howlin & Charman (2009); 4. Included in Rogers & Vismara (2008).

with pervasive developmental disorders and intellectual disabilities (Newsom & Hovanitz, 1997).

Adaptive behavior. The VABS (Sparrow et al., 1984) was the only measure for adaptive behavior used in the studies included in our analysis. The VABS yields standard scores on four domains: communication; daily living skills; socialization; and, for children younger than 6 years old, motor skills. Based on these scores it will also yield a standardized ABC. In our study we only used this composite score as we did not have access to the domain scores for most of the children. The VABS is widely regarded as the best available instrument for assessing adaptive behavior in children with autism (Newsom & Hovanitz, 1997).

Tests of Homogeneity and Publication Bias

Data were entered into the Comprehensive Meta-Analysis Software (Borenstein, Hedges, Higgins, & Rothstein, 2005). To determine whether all studies were drawn from a population of studies with a common mean effect size, we performed a test of homogeneity using the Q-statistic and I^2, utilizing these options in the Comprehensive Meta-Analysis Software. These tests were conducted for the whole group of nine studies together. The I^2 gives the proportion of the variance that be explained by between-study variance. Using the software, we also assessed potential publication bias by a funnel plot of the standard error and effect size for each study (Egger, Smith, Schneider, & Minder, 1997) and the trim and fill method (Duval & Tweedie, 2000) for both IQ and ABC outcomes.

Effect Size Measures

The standardized mean difference effects size for EIBI were computed for IQ and ABC using the same software. Effect sizes were computed for each study separately, but we also computed an overall effect size against the comparison and control groups. We used the Hedges's g effect size measure (Hedges & Olkin, 1985) to adjust for the relatively small sample sizes in the studies, typically less than 20 in each group. When computing an overall (meta-analytic) effect size the individual studies were weighted using the inverse of the variance, as is widely considered to be the best practice (Borenstein, Hedges, Higgins, & Rothstein, 2009).

RESULTS

Homogeneity

The Q-statistic was not statistically significant for either IQ, $Q(9) = 10.07$, $p = .345$, or ABC, $Q(7) = 8.50$, $p = .291$, scores across the nine identified studies. This indicated that all of the studies could be combined for one common effect size. We also calculated the between-study variance for IQ ($I^2 = 10.66$) and for ABC scores ($I^2 = 17.65$), and these data supported the homogeneity conclusion in that relatively small proportions of variance were explained by between-study variance. Given these findings, we used a fixed effects model for computing all effect sizes.

Effect Size Measures

The standardized mean difference effect size was calculated for IQ and ABC. The Howard, Sparkman, Cohen, Green, and Stanislaw (2005) study contributed both a control and a comparison group. We decided to calculate separate effect sizes for these. Hence, the total number of effect sizes for IQ was 10 from the nine studies included. Four studies had a comparison group and six studies had a control group. For ABC, four studies had a comparison group and four studies had a control group, the Howard et al. study again contributing one to each group, making the total number of effect sizes eight from the seven studies included. A forest plot of the effect sizes for each study and an overall effect size for IQ and ABC are shown in Figure 2.

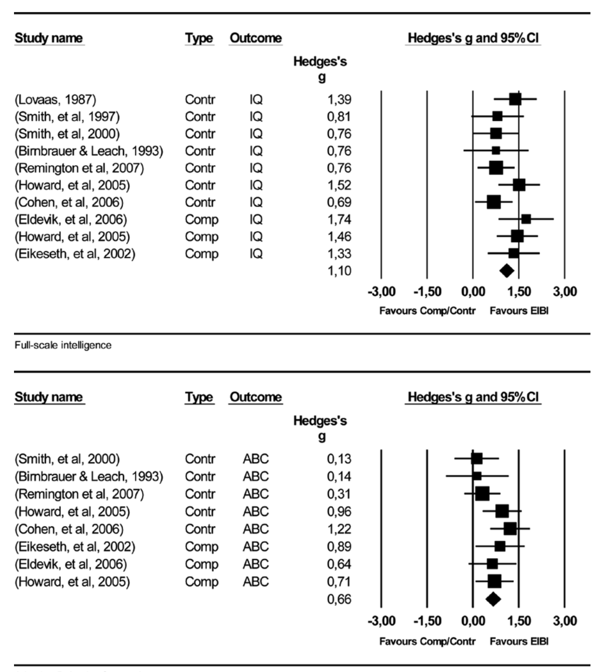

FIGURE 2 Forrest plots of standardized mean difference effect sizes (Hedges's g) and 95% confidence intervals (CIs). In the first plot effect sizes for full-scale IQ are shown and in the second ditto for adaptive behavior composites. Studies are grouped as either comparison or control. The fixed model effect size is computed against both the comparison and control studies and also an overall effect size is computed. EIBI = Early Intensive Behavioral Intervention.

The overall effect size for IQ change was 1.103 (95% confidence interval [CI] = .871, 1.335). The overall effect size for change in adaptive behavior composite scores was .660 (95% CI = .41, .90).

Publication Bias

We found no statistical or visual evidence of publication bias. Funnel plots of the standard error against effect sizes for IQ and ABC changes are shown in Figure 3. The Duval and Tweedie (2000) trim and fill method did not suggest the potential absence of any studies. However, the limitations of these techniques, particularly when there are few studies, means we cannot exclude publication bias.

DISCUSSION

Following EIBI treatment, our meta-analysis found an average large effect size for IQ change (based on 10 comparisons) and an average medium effect size for ABC change (based on 8 comparisons; Cohen, 1992). These estimates of effect size were also statistically significant from zero (the 95% CIs did not contain zero). We also found that the sample of studies was reasonably homogenous and that there was no evidence of publication bias.

The only other published meta-analysis we have found (Reichow & Wolery, 2009) reported an effect size of .69 for IQ change. The effect size for IQ change is thus somewhat higher in our analysis. There may be several explanations for the difference between our results

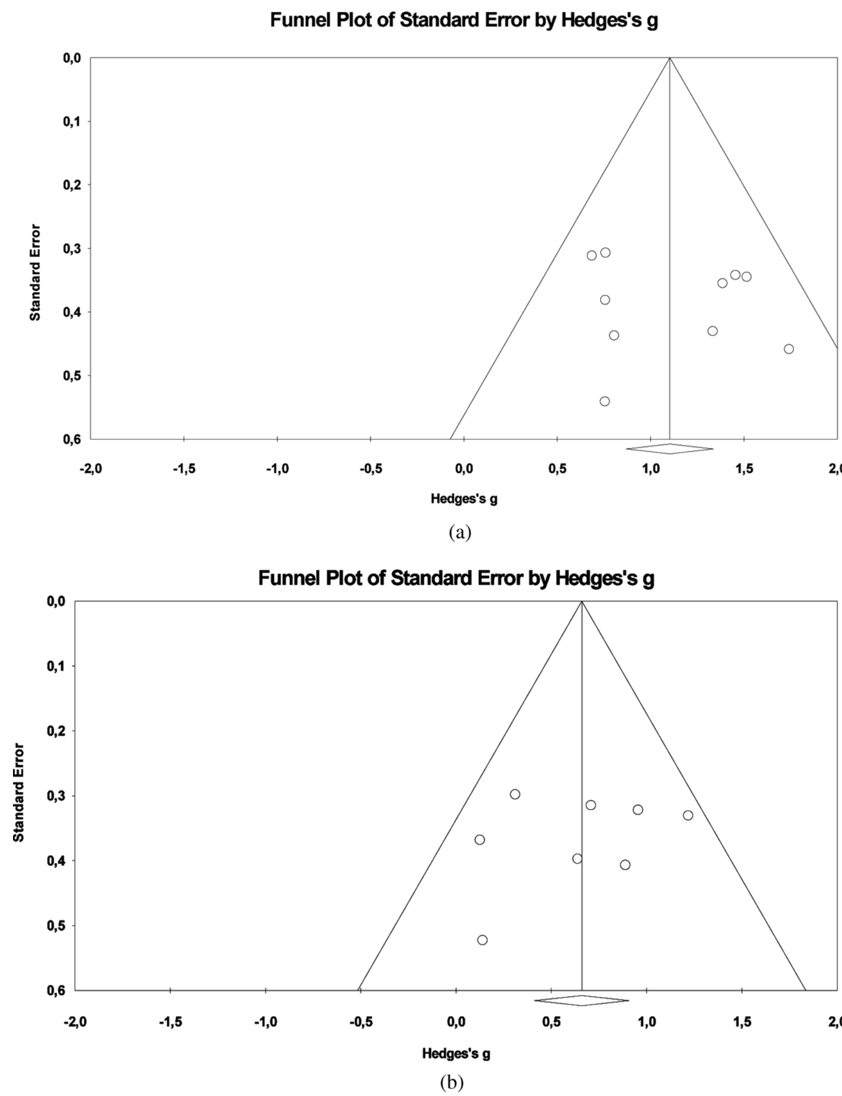

FIGURE 3 Funnel plots of effect sizes against the standard errors. In the first plot the full-scale IQ data are shown, and in the second plot the adaptive behavior composites are shown. The circles represent the studies included in the analysis, and the diamond represents the average effect size with a 95% confidence interval.

and those of Reichow and Wolery that relate to the steps we took to expand and improve on their analysis. First, to be able to use the more methodologically rigorous standardized mean difference effect size, we only included studies that had a control or a comparison group. In contrast, Reichow and Wolery computed their mean effect size based on the change within the EIBI group only. Second, we applied a more precise yet inclusive definition of EIBI, introduced by Green et al., (2002), that seems to us to be more in keeping with how other EIBI professionals define their field (e.g., Eikeseth, 2009). Reichow and Wolery employed more restricted criteria for including studies in their review, in that it had to be a replication of Lovaas's UCLA/YAP model and/or based on their treatment manuals. This may be the reason why the Howard et al., (2005) and the Remington et al., (2007) studies were not included in their analysis. However, in our opinion, the failure to include these two studies represents an inconsistency in the selection process. To us, the treatment provided in these studies is as much EIBI as the treatment provided in the studies that were included by Reichow and Wolery. Along the same lines, it seems inappropriate to us to include in the calculation of effect size a study that compared two models of EIBI service delivery (Sallows & Graupner, 2005). It is quite clear from the description of the provisions in this study that, although the groups differ in some respects, they are both examples of EIBI.

A third difference in our analysis is that we only included studies that reported full-scale intelligence scores. Reichow and Wolery did not make a distinction between performance based and full-scale intelligence measures. As we have noted earlier, this may skew results in either direction, especially when tests are used interchangeably over time. Fourth, our analysis was based on individual raw data from each study rather than the data reported in the published papers. This meant that we had a slightly different sample of children, even from the studies in common to both analyses (see Table 1).

As an extension to Reichow and Wolery's meta-analysis, we were able to include an analysis of another important outcome measure, namely, the adaptive behavior composite. This measure adds substantial validity to the outcomes, because it tells us more about the children's skills in daily life. Of interest, effects sizes were lower than for IQ. We also tested if intensity of treatment in itself may account for differences in outcome. This was possible by employing stricter criteria for what should constitute a comparison group. In our study, the comparison groups had to be given a provision of similar intensity (measured as weekly hours of 1:1 provision) as the EIBI groups. In the studies included here, this meant an "eclectic" provision. Although we agree with Reichow and Wolery's (2009) point that it is still hard to determine whether this means a specific common provision, we think it is valuable to treat them as a group, especially as eclectic provision is probably similar to a treatment as usual for many children with autism. In eclectic programs, the particular composition of treatments is to be adjusted to the individual child's needs and may thus vary a great deal across children and across time for a given child. Attempts to measure this have been made (e.g., Eikeseth et al., 2002), but it proved difficult for teachers in the eclectic groups (and thus for the researchers) to say what specific treatment they were using because they tended to blend and apply them depending on the child's behavior and needs through the day. Although difficult to specify, the eclectic approach seems to be the most common provision offered to children with ASD in service settings currently, even among those clinicians with behavior analysis training (Schreck & Mazur, 2008). Our results add to the serious concerns raised by Rogers and Vismara (2008) about eclectic treatment models.

Although we were able to refine Reichow and Wolery's meta-analysis, there are some serious limitations that remain, such that any conclusions need to be drawn with caution and to be considered tentative. First, the number of studies included in our analysis may be considered small, although it is above the median for reviews listed in the the Cochrane Database of Systematic Reviews. This database currently includes more than 3,000 reviews, and the median number of studies in a review is six (Borenstein et al., 2009). Second, a more serious limitation is the quality of the studies on effects of EIBI. Because of the lack of random assignment, only one study included in the present analysis met Type 1/highest level criteria of methodological rigor (Nathan & Gorman, 2002). Furthermore, the literature lacks comparisons between EIBI and other approaches, perhaps other than the eclectic one. Third, although there is a clear difference in outcome between EIBI and the comparison intervention, it should be noted that this may be due to differences in the amount and frequency of supervision and training. We did not have enough data to control for this in the present study. However, based on the information in the studies included, it is clear that the EIBI group in general received more frequent and more total hours of supervision and training. This remains a threat for the validity of conclusions about the superiority of EIBI in relation to comparison intervention. Fourth, we decided to include two effect sizes from the Howard et al. (2005) study, one for EIBI against the comparison group and one for EIBI against the control group. This is problematic because they are not independent of each other as both involve contrasts with a single EIBI group. We did all calculation only including the effect size from the comparison group and this did not alter the overall results in any significant manner; Hedges's $g = 1.048$ (95% CI = .80, 1.30) for IQ, and Hedges's $g = .607$ (95% CI = .34, .87) for ABC. Fifth, because of the limited number of studies and available variables we decided not to conduct an analysis of moderator variables that may explain variation in intervention outcome. However, this is certainly a priority for the future when more studies are published and more potential moderator variables can be analyzed.

Implications for Research, Policy, and Practice

With these limitations in mind, our general conclusions are very similar to those of other recent reviews: EIBI produces large to moderate effect sizes for changes in IQ and ABC scores for children with ASD when compared with no intervention controls and eclectic provision. These results support the clinical implication that EIBI at present should be an intervention of choice for children with ASD. However, randomized controlled trials comparing EIBI to other interventions are still needed. In particular, studies are needed where the comparison intervention is of similar intensity and where staff receive similar training and supervision.

REFERENCES

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 & Treatment of Children*, *10*, 352–366.

Bayley, N. (1969). *Bayley Scales of Infant Development*. New York: Psychological Corporation.

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

Beglinger, L., & Smith, T. (2005). Concurrent validity of social subtype and IQ after Early Intensive Behavioral Intervention in children with autism: A preliminary investigation. *Journal of Autism and Developmental Disorders*, *35*, 295–303.

Ben-Itzchak, E., & Zachor, D. A. (2007). The effects of intellectual functioning and autism severity on outcome of early behavioral intervention for children with autism. *Research in Developmental Disabilities*, *28*, 287–303.

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": Erratum. *Research in Developmental Disabilities*, *23*, 79–104.

Birnbrauer, J. S., & Leach, D. J. (1993). The Murdoch Early Intervention Program after 2 years. *Behaviour Change*, *10*, 63–74.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta analysis Version 2*. Englewood, NJ: Biostat.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. (2009). *Introduction to meta-analysis*. Hoboken, NJ: Wiley.

Butter, E. M., Mulick, J. A., & Metz, B. (2006). Eight case reports of learning recovery in children with pervasive developmental disorders after early intervention. *Behavioral Interventions*, *21*, 227–243.

Chambless, D. L., Baker, M. J., Baucom, D. H., Beutler, L. E., Calhoun, K. S., & Crits-Christoph, P. (1998). Update on empirically validated therapies, II. *The Clinical Psychologist*, *51*, 3–16.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, *6*, 7–18.

Chambless, D. L., Sanderson, W. C., Shoham, V., Bennett Johnson, S., Pope, K. S., & Crits-Christoph, P. (1996). An update on empirically validated therapies. *The Clinical Psychologist*, *49*, 5–18.

Cohen, H., Amerine-Dickens, M., & Smith, T. (2006). Early Intensive Behavioral Treatment: Replication of the UCLA Model in a community setting. *Developmental and Behavioral Pediatrics*, *27*, 145–155.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

Drew, A., Baird, G., Baron-Cohen, S., Cox, A., Slonims, V., Wheelwright, S., et al. (2002). A pilot randomised control trial of a parent training intervention for pre-school children with autism: Preliminary findings and methodological challenges. *European Child & Adolescent Psychiatry*, *11*, 266–272.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463.

Egger, M., Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple graphical test. *British Medical Journal*, *315*, 629–634.

Eikeseth, S. (2009). Outcome of comprehensive psycho-educational interventions for young children with autism. *Research in Developmental Disabilities*, *30*, 158–178.

Eikeseth, S., Smith, T., Jahr, E., & Eldevik, S. (2002). Intensive behavioral treatment at school for 4- to 7-year-old children with autism: 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.

Eldevik, S., Eikeseth, S., Jahr, E., & Smith, T. (2006). Effects of low-intensity behavioral treatment for children with autism and mental retardation. *Journal of Autism & Developmental Disorders*, *36*, 211–224.

Fenske, E. C., Zalenski, S., Krantz, P. J., & McClannahan, L. E. (1985). Age at intervention and treatment outcome for autistic children in a comprehensive intervention program. *Analysis & Intervention in Developmental Disabilities*, *5*, 49–58.

Gabriels, R. L., Hill, D. E., Pierce, R. A., Rogers, S. J., & Wehner, B. (2001). Predictors of treatment outcome in young children with autism. *Autism*, *5*, 407–429.

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

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

Harris, S. L., Handleman, J. S., Gordon, R., Kristoff, B., & Fuentes, F. (1991). Changes in cognitive and language functioning of preschool children with autism. *Journal of Autism & Developmental Disorders*, *21*, 281–290.

Hayward, D. W., Eikeseth, S., Gale, C., & Morgan, S. (in press). Assessing progress during treatment for young children with autism receiving intensive behavioural interventions. *Autism*.

Hedges, L., & Olkin, I. (1985). *Statistical models for meta-analysis*. New York: Academic Press.

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.

Howlin, P., Magiati, I., & Charman, T. (2009). A systematic review of early intensive behavioural interventions (EIBI) for children with autism. *American Journal on Intellectual and Developmental Disabilities*, *114*, 23–41.

Ingersoll, B., Schreibman, L., & Stahmer, A. (2001). Brief report: Differential treatment outcomes for children with Autistic Spectrum Disorder based on level of peer social avoidance. *Journal of Autism and Developmental Disorders*, *31*, 343–349.

Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., & Pickles, A. (2006). Autism from 2 to 9 years of age. *Archives of General Psychiatry*, *63*, 694–701.

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

Luiselli, J. K., Cannon, B. O. M., Ellis, J. T., & Sisson, R. W. (2000). Home-based behavioral intervention for young children with autism/pervasive developmental disorder: A preliminary evaluation of outcome in relation to child age and intensity of service delivery. *Autism: The International Journal of Research and Practice*, *4*, 426–438.

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

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.

Morris, S. B. (2000). Distribution of the standardized mean change effect size for meta-analysis on repeated measures. *The British Journal of Mathematical and Statistical Psychology*, *53*, 17–29.

Nathan, P., & Gorman, J. M. (2002). *A guide to treatments that work*. New York: Oxford University Press.

Newsom, C., & Hovanitz, C. A. (1997). Autistic disorder. In *Assessment of childhood disorders* (3rd ed., pp. 408–452). New York: Guilford.

Reed, P., Osborne, L. A., & Corness, M. (2007a). Brief report: Relative effectiveness of different home-based behavioral approaches to early teaching intervention. *Journal of Autism and Developmental Disorders*, *37*, 1815–1821.

Reed, P., Osborne, L. A., & Corness, M. (2007b). The real-world effectiveness of early teaching interventions for children with autism spectrum disorder. *Exceptional Children*, *73*, 417–433.

Reichow, B., & Wolery, M. (in press). 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*.

Remington, B., Hastings, R. P., Kovshoff, H., Espinosa, F., Jahr, E., Brown, T., et al. (2007). Early intensive behavioral intervention: Outcomes for children with autism and their parents after two years. *American Journal on Mental Retardation*, *112*, 418–438.

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

Roid, G. M., & Miller, L. M. (1997). *Leiter International Performance Scale–Revised: Examiners manual*. Wooddale, IL: Stoelting.

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

Schreck, K. A., & Mazur, A. (2008). Behavior analyst use of and beliefs in treatments for people with autism. *Behavioral Interventions*, *23*, 201–212.

Sheinkopf, S. J., & Siegel, B. (1998). Home-based behavioral treatment of young children with autism. *Journal of Autism & 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., Eikeseth, S., Klevstrand, M., & Lovaas, O. I. (1997). Intensive behavioral treatment for preschoolers with severe mental retardation and pervasive developmental disorder. *American Journal on Mental Retardation*, *102*, 238–249.

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.

Solomon, R., Necheles, J., Ferch, C., & Bruckman, D. (2007). Pilot study of a parent training program for young children with autism: The PLAY Project Home Consultation program. *Autism*, *11*, 205–224.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (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: Outcome from the Children's Toddler School. *Journal of Positive Behavior Interventions*, *6*, 67–82.

Stutsman, R. (1948). *Guide for administering the Merrill-Palmer Scale of Mental Tests*. New York: Harcourt, Brace & World.

Thorndike, R. L., Hagen, E. R., & Sattler, J. M. (1986). *The Stanford-Binet intelligence scale* (4th ed.). Chicago: Riverside.

Wechsler, D. (1974). *The Wechsler Intelligence Test for Children–Revised*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1989). *Wechsler Preschool and Primary Scale of Intelligence–Revised*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1993). *Wechsler Intelligence Scale for Children* (3rd ed.). San Antonio, TX: Psychological Corporation.

Weiss, M. J. (1999). Differential rates of skill acquisition and outcomes of early intensive behavioral intervention for autism. *Behavioral Interventions*, *14*, 3–22.