Research in Autism Spectrum Disorders 51 (2018) 18–31

Contents lists available at ScienceDirect

Research in Autism Spectrum Disorders

journal homepage: www.elsevier.com/locate/rasd

The effectiveness of applied behavior analytic interventions for children with Autism Spectrum Disorder: A meta-analytic study

Maria K. Makrygianni*, Angeliki Gena, Sofia Katoudi, Petros Galanis

National and Kapodistrian University of Athens, Department of Philosophy, Pedagogy and Psychology, Panepistimiopolis, 15784 Illisia (3rd Floor), Athens, Greece

A R T I C L E I N F O

Number of reviews completed is 2

Keywords:
Autism Spectrum Disorder
Effectiveness
Applied behavior analysis
Meta-analysis

A B S T R A C T

Background: Behavior Analytic interventions are interventions based on the principles of Applied Behavior Analysis (ABA). They are widely used with children with Autism Spectrum Disorder (ASD), they are highly effective, and are the most-widely studied types of interventions for children with ASD. Nevertheless, findings are not consistent regarding the *degree* of its effectiveness. The purpose of this meta-analysis was to provide an up-to-date and thorough evaluation of ABA programs for children with ASD using state of the art meta-analytic methodology.
Method: For the purposes of the present analysis, 29 studies met the inclusion criteria and were consequently analyzed. Effectiveness was evaluated in terms of three domains pertaining to child-related variables: a) IQ scores provided by verbal and non-verbal standardized tests, b) receptive and expressive language, and c) adaptive behavior. This evaluation included comparisons of pre- and post-intervention outcomes rather than comparisons between experimental and control groups.
Results and conclusions: The findings indicated that ABA programs are moderately to highly effective bringing significant benefits for children with ASD in the aforementioned areas. Specifically, they were very effective in improving intellectual abilities ($g = 0.740$); moderately to very effective in improving communication skills ($g = 0.650$), *expressive-language skills* ($g = 0.742$) *and receptive- language skills* ($g = 0.597$); moderately effective in improving IQ provided by non-verbal tests ($g = 0.463$), adaptive behavior (in total) ($g = 0.422$), socialization ($g = 0.444$); and had low effectiveness in improving daily living skills ($g = 0.138$).

1. Introduction

The prevalence or diagnosis of Autism Spectrum Disorder (ASD) has increased dramatically as indicated by a great number of epidemiological studies throughout the world (e.g. South Korea, Kim et al., 2011; United Kingdom, Baird et al., 2006; Iceland, Saemundsen, Magnússon, Georgsdóttir, Egilsson, & Rafnsson, 2013). According to the Autism and Developmental Disabilities Monitoring Network in the USA, during the decade 2002–2012 the diagnosis of ASD was increased by 121%. The increasing numbers of people diagnosed with ASD, along with the severity of the condition, underline the importance of identifying most effective therapeutic approaches for its treatment.

Applied Behavior Analytic (ABA) interventions are recognized as the most effective evidence-based interventions for children with ASD (Fein et al., 2013). The term ABA intervention refers to treatment approaches that: (a) are implemented systematically following the principles of applied behavior analysis; (b) are applied as early as possible in the child's life, preferably before the age of

* Corresponding author.
E-mail addresses: mmakrigianni@yahoo.gr (M.K. Makrygianni), agena@ppp.uoa.gr (A. Gena), pgalanis@ppp.uoa.gr (P. Galanis).

https://doi.org/10.1016/j.rasd.2018.03.006
Received 6 June 2017; Received in revised form 1 March 2018; Accepted 21 March 2018

3 years; (c) are usually provided in a student-teacher ratio of one-to-one before generalization procedures are used; (c) are individualized, comprehensive, and target a great number of skills; (d) incorporate skills that are targeted following a hierarchy based on typical development; and (d) are used in conjunction with parent-education services (Healy & Lydon, 2013; Virues-Ortega, 2010). The origins of this approach are traced to the UCLA-Young Autism Project that Lovaas and his colleagues developed and run in the 1980′s. This model became very well known after the influential and controversial publication that demonstrated its powerful effects imposing dramatic improvements in children with ASD (Lovaas, 1987). The findings of this study constitute a milestone for the treatment of this potentially incapacitating disorder. Since then, a great number of studies have been carried out that provide additional support regarding the effectiveness of ABA intervention. This effectiveness translates into *increases or improvement*: in IQ scores, language skills, and adaptive behavior and at the same time into decreases: in autism-related symptomatology and/or symptom intensity, in the need for support during school inclusion, and in challenging behavior (e.g., Anderson, Avery, DiPietro, Edwards, & Christian, 1987; Birnbrauer & Leach, 1993; Eikeseth, Klintwall, Jahr, & Karlsson, 2012; Fava et al., 2011; McEachin, Smith, & Lovaas, 1993; Strauss et al., 2012; Weiss, 1999). Even though all ABA intervention studies measure behavior changes of young children with ASD as a result of the intervention applied, they vary from one another, a great deal, in terms of design parameters, population characteristics, characteristics of the intervention, and outcomes of the intervention which leads to the challenge of evaluating the parameters that may contribute most to the effectiveness of ABA interventions and the degree of their efficacy.

Despite the questions that arise from the aforementioned variability (e.g., Healy & Lydon, 2013; Howlin, Magiati, & Charman, 2009; Reichow, Barton, Boyd, & Hume, 2012; Warren et al., 2011), we may reach safe conclusions about treatment efficacy by using systematic qualitative and quantitative analysis of the findings of at least some of the intervention studies. Specifically, meta-analytic research provides an objective medium for the assessment of ABA intervention which facilitates summarizing, integrating, and interpreting a group of quantitative empirical studies with similar methodology. A meta-analysis helps us convert the results from different studies to a common metric, and statistically explore the relations between the characteristics and the findings of those studies. It is a research tool that was developed at the end of the 1970s (Lipsey & Wilson, 2001) and originally used in the social sciences. Since then, meta-analysis has become a widely accepted research tool used in a variety of disciplines. Despite its strengths, it is not always possible to use meta-analytic methodology to evaluate sets of data (Lipsey & Wilson, 2001).

The effectiveness of ABA intervention has already been the subject of nine meta-analytic studies (Eldevik et al., 2009, 2010; Kuppens & Onghena, 2012; Makrygianni & Reed, 2010; Peters-Scheffer, Didden, Korzilius, & Sturmey, 2011; Reichow et al., 2012; Reichow & Wolery, 2009; Spreckley & Boyd, 2009; Virues-Ortega, 2010) published in peer-reviewed journals. Even though each of those studies incorporated different inclusion criteria, they all demonstrated the effectiveness of ABA intervention in the treatment of ASD whether pre and post-intervention methodology or group designs were used to evaluate effectiveness.

The purpose of the present meta-analysis was to assess the efficacy of ABA intervention in improving IQ scores (provided by verbal and non-verbal standardized tests), receptive and expressive language, and adaptive behavior (total and subcategories) of children with ASD. A total of 29 studies were meta-analyzed, 20 of which were included in previous meta-analytic studies and 9 that were meta-analyzed for the first time (i.e., Eldevik, Hastings, Jahr, & Hughes, 2012; Fernell et al., 2011; Flanagan, Perry, & Freeman, 2012; Grindle et al., 2012; Peters-Scheffer, Didden, Mulders, & Korzilius, 2010; Peters-Scheffer, Didden, Mulders, & Korzilius, 2013; Rivard, Terroux, & Mercier, 2014; Smith, Klorman, & Mruzek, 2015; Zachor & Ben-Itzchak, 2010). There are multiple advantages to the present study: (a) It is the only meta-analytic study since 2012 that provides an assessment of the efficacy of ABA treatment by comparing the participants' performance before and after the application of treatment. (b) It meta-analyzed a greater number of studies than any prior meta-analysis of ABA intervention effectiveness. (c) It is one of a few meta-analytic studies that analyzed a great number of variables (eight variables) associated with treatment.

We anticipate that this study will contribute toward a thorough, comprehensive, and up-to-date assessment of the effectiveness of ABA intervention.

2. Methods

2.1. Search strategy and selection of studies

A thorough search of the literature was conducted from January 1987 to October 2017 in order to retrieve studies that met inclusion criteria for the present meta-analysis. This search was conducted by three of the authors. Computerized literature searches of PubMed and Science Direct were conducted using the keywords*: behavio(u)r analytic, or applied behavio(u)r analysis, or behavio(u)ral treatment or behavio(u)ral intervention* in combination with *autism, or autistic, or PDD.* Additionally, recent publications, reviews, and meta-analyses were also inspected manually (e.g., Eldevik et al., 2009; Kuppens & Onghena, 2012; Makrygianni & Reed, 2010; Reichow, 2012; Reichow & Wolery, 2009; Spreckley & Boyd, 2009; Virues-Ortega, 2010) and recommendations from experts in the field were taken into account in order to cross-check whether all the relevant published studies had been located.

The inclusion criteria for the ABA intervention studies in the present meta-analysis were the following: (a) using experimental or quasi-experimental design; (b) using ABA principles and teaching techniques, in general, rather than a specific ABA approach, such as pivotal response training or verbal behavior; (c) addressing various domains of the child's life rather than limiting the scope of the intervention to one area, such as language development or communication; (d) including a sample of children diagnosed with ASD, Autism, Autistic Disorder, PDD, or PDD-NOS; (d) providing child assessment measures on at least one of the following domains: intelligence (provided by verbal and non-verbal standardized tests), receptive language, expressive language, adaptive behavior (composite, communication, daily living skills or socialization), and; (e) using one or more of the following: verbal or non-verbal

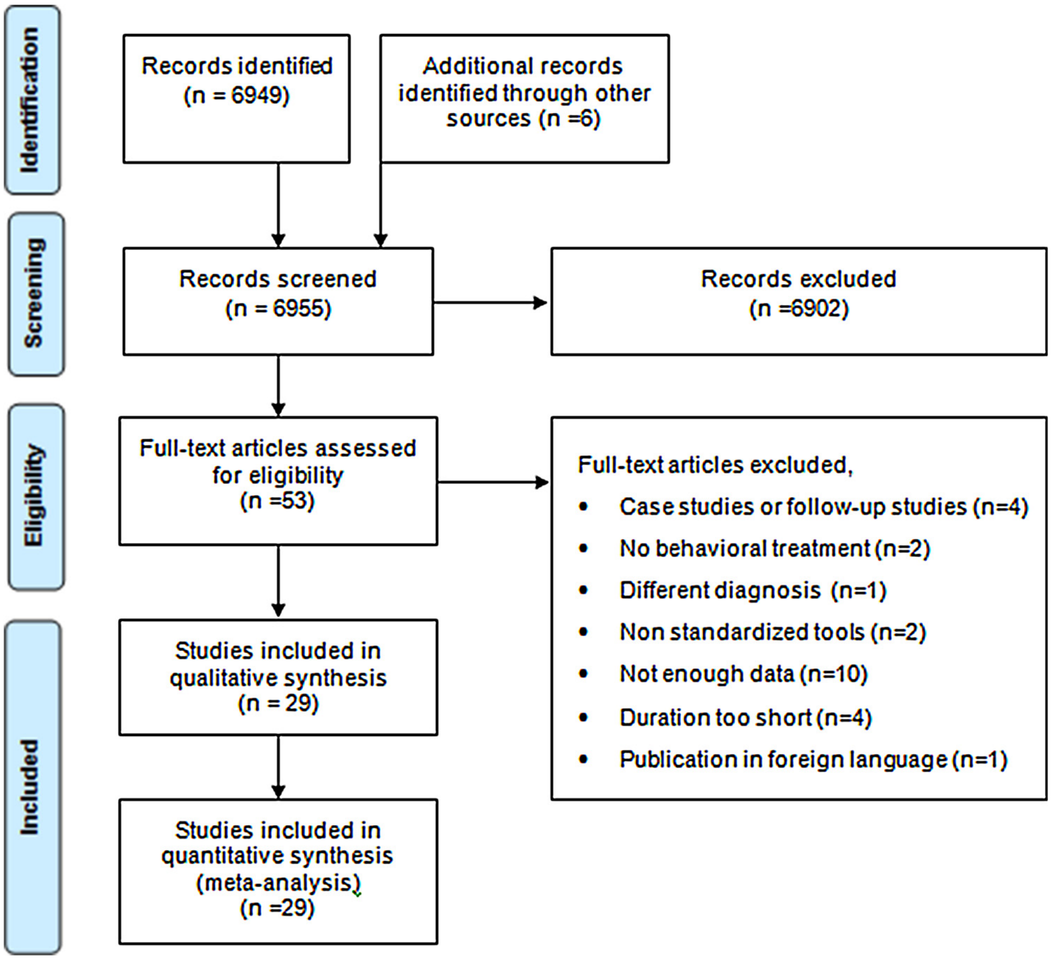

Fig. 1. The search and selection procedure (flow diagram).

standardized scales for the assessment of intelligence (i.e., IQ or ratio IQ of Bayley Scales of Infant Development, Cattell Infant Intelligence Scale, Mullen scales of Early Learning, Stanford–Binet Intelligence Scales, Wechsler Preschool and Primary Scale of Intelligence, Wechsler Intelligence Scale for Children, Leiter International Performance Scale, Merrill-Palmer Scales of Mental Tests), scores of the Reynell Developmental Language Scales for the assessment of receptive and expressive language, and the Vineland Adaptive Behavior Scales for the assessment of adaptive behavior (i.e., standard scores of one or more of the subscales or the composite); (f) providing both a pre-treatment and a post-treatment assessment; (g) having a minimum treatment duration of seven months; and (h) being published in English.

A total of 29 studies met the inclusion criteria and were consequently included in this meta-analysis (see Fig. 1). They were published between 1987 and 2015 and were conducted in seven different countries.

2.2. Meta-analysis

For the purposes of this meta-analysis, the random effects model was selected as most suitable since we anticipated high variability among the characteristics of the samples and the treatment methodologies used in the studies that were meta-analyzed. The random effects model helps us estimate the mean of a range of effects and not one “true” effect (as estimated by the fixed effect Model). Overall, this Model is considered to be more suitable for applied research than the fixed effect Model (Borenstein, Hedges, Higgins, & Rothstein, 2010).

The meta-analytic procedure began with the extraction of relevant data from each study, which was conducted by three independent raters to ensure a reliable outcome. Interobserver agreement across all measures, on the average, was 90%. Disagreements were discussed among raters and agreement was finally reached for all ratings after joined review of the points of disagreements. Quantitative analysis of the data was conducted using the Comprehensive Meta-Analysis (CMA) software, version 2. Statistical analysis was conducted on multiple levels: (a) to estimate mean effect sizes, (b) to identify publication bias, (c) to assess homogeneity.

2.2.1. Standardized mean change effect size

There are two crucial steps in meta-analysis: coding and conversion of the provided outcomes to a common metric – a process that potentiates comparisons among the studies that are meta-analyzed. Effect Size (ES) is one of the indexes of magnitude and direction of the treatment effect (Hedges & Olkin, 1985; Lipsey & Wilson, 2001; O’Mara, Marsh, & Craven, 2005). Specifically, ES constitutes a quantitative assessment of the magnitude and the power of a phenomenon (Kelley & Preacher, 2012). The type of ES, used in the present study, was the standardized mean change (ES_{change}) which expresses the difference between pre- and post-treatment measures. For the calculation of standardized mean change, *Hedges's g* was used because it constitutes a conservative estimate (Hedges, 1981). For the interpretation of ESs, Cohen provided "Rules-of-Thumb" suggesting that 0.2 represents a small ES, 0.5 represents a medium ES, and 0.8 represents a large ES (Lipsey & Wilson, 2001).

The resent study assessed the effectiveness of ABA intervention on three developmental domains: (a) cognitive abilities provided by verbal and non-verbal standardized tests; (b) language abilities: receptive and expressive language; (c) adaptive behavior: communication, socialization, daily living skills, and overall adaptive behavior. Consequently, in total, there were eight variables, and for each one of them a standardized mean change ES was calculated (see inclusion criteria for operational definitions of those variables).

ESs for each of the eight variables were checked for extreme values (outliers). Only one extreme value was spotted (Cohen, Amerine-Dickens, & Smith, 2006 on receptive language) and it was recoded to a more moderate value (mean +2 Standard Deviations). The identification and proper handling of outliers is important in avoiding distortion of the results (Lipsey & Wilson, 2001; Rosenthal, 1995; Viechtbauer & Cheung, 2010). According to Lipsey and Wilson (2001), outliers can be handled either through deletion or through adjustment in less extreme values. In the present study, the second option was chosen.

2.2.2. Publication bias

Publication bias refers to the possibility that published studies, in general, are systematically unrepresentative of the population of conducted studies because of a tendency to accept for publication studies that demonstrate significant effects and to reject studies with insignificant or inconclusive findings (Rothstein, Sutton, & Borenstein, 2006). Publication bias constitutes a major threat to research validity and meta-analytic research is particularly vulnerable to this risk. Despite this risk, meta-analytic research is well justified (Rothstein et al., 2006). Publication bias, in the present study was assessed using three procedures: visual inspection of *funnel plots, the Egger's test* (Egger, Smith, Schneider, & Minder, 1997), and the *Begg and Mazumdar rank correlation test* (Begg & Mazumdar, 1994). Asymmetry in *funnel plot* reflects possible publication bias (Sterne & Egger, 2001). The Egger's test, which is based on a simple linear regression analysis, quantifies the bias captured by the funnel plot. Finally, the Begg and Mazumdar (1994) *non-parametric test* assesses for publication bias, using the rank correlation (Kendall's tau) between intervention effect estimates and their variances.

2.2.3. Homogeneity test

The homogeneity of the data in meta-analysis was examined using two statistical tests: the *Cochran's Q statistic* (or chi-square test) and the *I^2 statistics*. The *Cochran's Q statistic* evaluates whether the variability of the ESs is larger than expected on the basis of Standard Errors (SEs) (heterogeneity) or is at the expected level (homogeneity). A significant *Q-statistic* indicates significant heterogeneity among the ESs and may reflect the interfering of confounding variables (Lipsey & Wilson, 2001). Nevertheless, there has been criticism for the use of *Q statistic* because it solely informs about the presence or absence of heterogeneity, but not how extensive it actually is.

Apart from *Q statistics,* the *I^2 statistics*, that we used, indicate the percentage of the total variability in a set of ESs that is attributable to true heterogeneity (Higgins & Thompson, 2002; Huedo-Medina, Sánchez-Meca, Marín-Martínez, & Botella, 2006). A rough rule for interpreting *I^2 statistics* is the following: (a) for I^2 = 0–40% heterogeneity is low, (b) for I^2 = 30–60% heterogeneity is moderate, (c) for I^2 = 50–90% heterogeneity is substantial and (d) for I^2 = 75–100% heterogeneity is considerable (Higgins & Green, 2011).

3. Results

3.1. Study characteristics

The main characteristics and findings of the studies included at the present meta-analysis are shown in Tables 1 and 2. The total number of children, included in the experimental groups of the 29 studies under review, was 831 and had a diagnosis of either ASD or PDD. Out of those studies, 22 provided information about boys to girls' ratio, 21 provided IQ scores obtained from verbal standardized tests and 10 IQ scores obtained from non-verbal standardized tests, 17 provided Vineland Adaptive Behavior Scales (VABS) composite scores and 7 receptive and expressive language scores, derived mainly from the Reynell Developmental Language Scales. Supplementary measures were reported for various aspects related to treatment (i.e., play, academic, self-help, visual-spatial, and motor skills, language, imitation, social and emotional development, learning rate, stereotypic behaviors, behavior problems, and severity of ASD symptoms) in 22 of the reviewed studies.

The participants' mean age at the beginning of the study ranged from 25.10 to 66.31 months (mean 41.89 months). The boys-to-girls ratio was 5.1:1. Mean standard scores at the beginning of the study were 54.46 for Full Scale IQ (provided by verbal tests) and 74.47 for IQ provided by non-verbal tests. The mean composite standard score on the VABS was 60.64. The mean standard scores for receptive and expressive language were 46.22 and 47.10 respectively.

Interventions were provided for 25.83 h per week on the average (range: 4.98–40.00 h/week), with mean duration of treatment 23.57 months (range: 8.00–93.00 months). In 18 studies the intervention was organized and directed by a specialized center, in 8 studies by community services, in 2 studies by universities, while in 1 study the intervention was organized and directed by parents. Treatment was provided in a single setting in 18 studies, while in the rest of them it was provided in multiple sites. Participants' homes were reported as intervention settings in 19 studies, school in 13 studies, community settings in 7, and a specialized center in 6 studies. Twenty one studies reported provision of parent training to prepare parents to undertake therapeutic responsibilities.

Table 1
Characteristics of the included studies.

Study	Participants[a]	Pre-treatment test scores[b]				ABA Intervention			
		IQ (verbal test)	IQ (Non-verbal tests)	Adaptive Behavior[c]	Language	Settings	h/w	Duration (months)	Parent training
Anderson et al. (1987) (USA)	N: 13 [10 boys] CA: 44.30 (10.80)	55.25 (19.64)	–	–	–	Home	15.00 – 25.00	12.00	Yes
Ben-Itzchak and Zachor (2007) (Israel)	N: 25 [23] CA: 26.60 (20–32)	70.67 (17.01)	–	–	–	Center	35.00	12.00	Yes
Bibby et al. (2002) (UK)	N: 22 [–] CA: 43.36 (12.60)	50.81 (20.60)	–	54.50 (13.00)	–	Home	30.30	31.60	No[d]
Cohen et al. (2006) (USA)	N: 21 [18] CA: 30.20 (5.80)[e]	61.60 (16.40)	82.40	69.80	Rec: 51.70 (15.20) Exp: 52.90 (14.50)	Home, school, community	35.00 – 40.00	36.00	Yes
Eikeseth, Smith, Jahr, and Eldevik (2002) (Norway)	N: 13 [8] CA: 66.31 (11.31)	61.92 (11.31)	77.54 (30.21)	55.77 (8.96)	Rec: 49.03 (16.42) Exp: 45.12 (13.44)	School	28.00	12.00	Yes
Eikeseth et al. (2012) (Norway)	N: 35 [29] CA: 46.80 (10.80)	–	–	67.00 (10.30)	–	School, home	23.00	12.00	No
Eldevik, Eikeseth, Jahr, and Smith (2006) (Norway)	N: 13 [10] CA: 53.00 (9.50)	41.00 (15.20)	68.20 (28.30)	52.50 (3.90)	Rec: 37.30 (11.70) Exp: 33.80 (10.60)	School	12.50	20.30	Yes
Eldevik et al. (2012) (Norway)	N: 31 [25] CA: 42.20 (9.00)	51.60 (16.90)	–	62.50 (8.20)	–	School	13.60	25.10	No
Fernell et al. (2011) (Sweden)	N: 93 [–] CA: 37.60 (8.50)	–	–	67.90 (9.88)	–	School, Home	15.00 – 40.00	20.90	No
Flanagan et al. (2012) (Canada)	N: 61 [53] CA: 42.93 (11.53)	–	–	55.38 (7.00)	–	Center	25.81	27.84	No
Grindle et al. (2012) (UK)	N: 9 [–] CA: 58.2 (43–68)	61.43 (19.18)	–	58.25 (6.59)	–	School	16.00	24.00	Yes
Harris, Handleman, Gordon, Kristoff, and Fuentes (1991) (USA)	N: 9 [8] CA: 50.11 (40–62)	67.56 (16.16)	–	–	–	Center, Home	–	10.89	No
Harris and Handleman (2000) (USA)	N: 27 [23] CA: 49.00 (31–65)	59.33 (24.20)	–	–	–	Center, Home	35.00 – 45.00	93.00	Yes
Hayward, Eikeseth, Gale, and Morgan (2009) (UK)	N: 44 [34] CA: 35.00 (6.00)	53.80 (14.00)	75.50 (20.03)	63.50 (8.80)	Rec: 57.17(–) Exp: 56.13(–)	Home, community	35.80	13.10	Yes
Howard, Sparkman, Cohen, Green, and Stanislaw (2005) (USA)	N: 29 [25] CA: 30.86 (5.16)	58.54 (18.15)	80.14 (11.86)	70.46 (11.85)	Rec: 52.16 (18.44) Exp: 51.88 (12.91)	Home, school, community	25.00 – 40.00	14.21	Yes
Lovaas (1987) (USA)	N: 19 [16] CA: 34.60 (–)	53.00 (–)	–	–	–	Home, School, Community	40.00	24.00	Yes
Magiati et al. (2007) (UK)	N: 28 [27] CA: 38.00 (7.2)	–	83.00 (27.90)	59.60 (6.20)	–	Home	32.40	25.50	Yes
Peters-Scheffer et al. (2010) (Netherland)	N: 12 [–] CA: 53.50 (5.52)	47.00 (10.33)	–		–	School	6.29	8.00	Yes
Peters-Scheffer et al. (2013) (Netherland)	N: 20 [18] CA: 62.52 (16.96)[f]	40.66 (20.07)			–	School	4.98	24.00	Yes
Reed, Osborne, and Corness (2007) (UK)	N: 12 [11] CA: 40.00 (32–47)	–	–	58.2 (8.00)	–	Home	30.40	9.00	Yes
Remington et al. (2007) (UK)	N: 23 [–] CA: 35.70 (4.0)	61.43 (16.43)		–	–	Home	25.60	24.00	Yes
Rivard et al. (2014) (Canada)	N: 93 [70] CA: 46 (0.57)	60.1 (16.4)	70.2 (18.5)	–	–	Center	16.00–20.00	12.00	Yes
Sallows and Graupner (2005) (UK)	N: 13 [11] CA: 33.23 (3.89)	50.85 (10.57)	70.58 (16.54)	59.54 (5.31)	Rec: 38.85 (6.09) Exp: 47.92 (6.17)	Home, school, community	37.57	48.00	Yes

(continued on next page)

Table 1 (*continued*)

Study	Participants[a]	Pre-treatment test scores[b]				ABA Intervention			
		IQ (verbal test)	IQ (Non-verbal tests)	Adaptive Behavior[c]	Language	Settings	h/w	Duration (months)	Parent training
Sheinkopf and Siegal (1998) (USA)	N: 4[g] [−] CA: 36.75 (5.62)	–	77.25 (15.34)	–	–	Home	19.45	15.73	Yes
Smith, Eikeseth, Klevstrand, and Lovaas (1997) (USA/Norway)	N: 11 [11] CA: 36.00 (6.90)	27.82 (4.90)	–	–	–	Home, community	30.00	35.00	Yes
Smith, Groen, and Wynn (2000) (USA)	N: 15 [12] CA: 36.07 (6.00)	50.53 (11.18)	59.88 (−)	63.44 (9.35)	Rec: 37.34(−) Exp: 41.94(−)	Home, school, community	24.52	33.44	Yes
Smith et al. (2015) (USA)	N: 71 [60] CA: 39.24 (7.8)	58.80 (13.39)	–	62.68 (9.02)	–	Center	16.61	24.00	No
Weiss (1999) (USA)	N: 20 [19] CA: 41.50 (20–65)	–	–	49.85 (7.84)	–	Home	40.00	24.00	Yes
Zachor and Ben Itzchak (2010) (Israel)	N: 45 [−] CA: 25.10 (3.9)	–	–	–[h]	–	Home	20.00	12.00	No

[a] N: Number of participants [N of boys], CA: Chronological Age at the beginning of the study in months.
[b] Mean scores (Standard Deviation or Range).
[c] Vineland Adaptive Behavior Scales Composite Score.
[d] No parent training was conducted or not indicated by the authors.
[e] Age at diagnosis.
[f] Mean CA of the participants in both groups (experimental and control).
[g] The authors provided the mean of pre and post-treatment IQ scores (62.8 and 89.7 respectively) from 9 of the participants. However these scores derived from various verbal and non-verbal pre and post tests and do not represent a full scale IQ. Only 4 of the participants received the same type (non-verbal) of standardized test prior and after the intervention.
[h] Only VABS subscales scores were provided. Communication: 67, Daily living skills: 67.70, Socialization: 67.80.

Table 2
Summary of outcomes.

Study	Change scores				Other positive outcomes[a]
	IQ (verbal test)	IQ (Non verbal tests)	Adaptive Behavior	Language	
Anderson et al. (1987)	4.12				Improvement in language, social, academic, and self-help skills. Accelerated learning rate. Partial school inclusion in reg. schools: 23%
Ben-Itzchak and Zachor (2007)	17.23	–	–	–	Improvement in imitation, non-verbal communication, language and play skills. Reduction of stereotypic behavior
Bibby et al. (2002)	4.14	–	8.90	–	Reduction of the severity of ASD. School inclusion in reg. schools with support: 34.87%
Cohen et al. (2006)	25.40	13.00	9.00	Rec: 20.30 Exp: 25.10	School inclusion in reg. schools without support: 28.57%, with support: 52.38%
Eikeseth et al. (2002)	17.16	17.46	11.23	Rec: 9.44 Exp: 22.27	
Eikeseth et al. (2012)	–	–	8.30	–	Reduction of the severity of ASD
Eldevik et al. (2006)	8.20	8.70	−0.10	Rec: 7.00 Exp: 11.00	Reduction of the severity of ASD
Eldevik et al. (2012)	15.00	–	5.90	–	
Fernell et al. (2011)	–	–	0.31	–	Reduction of the severity of ASD
Flanagan et al. (2012)	–	–	0.96	–	Reduction of the severity of ASD
Grindle et al. (2012)	8.43	–	10.25	–	Improvement in language, social, play, academic, self-help and motor skills
Harris and Handleman (2000)	18.26	–	–	–	Improvement in language. School inclusion in reg. schools without support: 11.11%, with support: 29.62%
Harris et al. (1991)	18.77	–	–	–	
Hayward et al. (2009)	16.10	10.40	6.40	Rec: 0.83 Exp: 2.44	
Howard et al. (2005)	31.34	21.53	10.86	Rec: 19.15 Exp: 18.58	Accelerated learning rate
Lovaas (1987)	30.30	–	–	–	School inclusion in reg. schools without support: 47.36%
Magiati et al. (2007)	–	−4.60	−2.10	–	Improvement in language and play skills. Reduction of the severity of ASD. School inclusion in reg. schools with support: 82.14%
Peters-Scheffer et al. (2010)	8.83	–	–	–	
Peters-Scheffer et al. (2013)	7.46	–	–	–	Improvement in language, play and interpersonal skills. Reduction of the severity of ASD and behavior problems
Reed et al. (2007)	–	–	2.9	–	Improvement in educational functioning, imitation, perception, fine and gross motor skills, eye-hand coordination, nonverbal and verbal conceptual ability. Reduction of the severity of ASD and behavior problems.
Remington et al. (2007)	12.05	–	–	–	Improvement in joint attention. Partial school inclusion in reg. schools: 73.91%
Rivard et al. (2014)	6.1	7.1	–	–	Reduction of the severity of ASD. Improvement in adaptive behavior and socioaffective competency
Sallows and Graupner (2005)	22.23	7.00	9.46	Rec: 17.00 Exp: 5.46	Reduction of the severity of ASD. School inclusion in reg. schools without support: 61.53%, with support: 23.07%
Sheinkopf and Siegal (1998)	–	19.50	–	–	Reduction of the severity of ASD. School inclusion in reg. schools without support: 27.27%, with support: 18.18%
Smith et al. (1997)	8.00	–	–	–	Improvement in language. Reduction of behavior problems
Smith et al. (2000)	15.96	8.50	−2.25	Rec: 8.23 Exp: 5.39	Improvement in academic skills. School inclusion in reg. schools without support: 26.66%
Smith et al. (2015)	6.13	–	−2.79	–	Reduction of the severity of ASD
Weiss (1999)	–	–	33.75	–	Reduction of the severity of ASD. School inclusion in reg. schools without support: 35.00%, with support: 8 (40%)
Zachor and Ben Itzchak (2010)	–	–	–[b]	–	Improvement in language

[a] Other positive results reported by the studies for all or at least some of the participants, including the percentage of participants included in regular education classes with or without support.
[b] Improvement in communication and socialization 5.9 and 1.8 points respectively. No change on daily living skills.

Seven studies used quasi-experimental, within-subjects, pre-post design to evaluate the efficacy of ABA interventions on ASD. The remaining studies used a quasi-experimental between-groups pre-post design, comparing the performance of an experimental group, receiving ABA intervention, with a control group, receiving an eclectic or "treatment-as-usual" intervention. Only two studies used a random experimental between-groups pre-post design.

Mean standard scores after ABA intervention were 68.80 for Full Scale IQ obtained from verbal tests (mean change of scores between intake and follow up: 14.34) and 85.33 for IQ scores obtained from non-verbal standardized tests (mean change of scores: 10.86). Mean composite standard scores at post-test on the VABS was 67.17 (mean change of scores: 6.53), for receptive language was 57.93 (mean change of scores: 11.71) and for expressive language was 59.99 (mean change of scores: 12.89). Thirteen studies reported reduction of the severity of ASD symptomatology. Twenty-four percent (24%) of the participants in 3 studies that provided the relevant information were reclassified either as non-autistic or received a diagnosis for a milder type of ASD. Other positive outcomes reported by the studies included reduction of stereotypic behavior and behavior problems, improvement in joint attention, play skills, self-help, visual-spatial and motor skills, imitation, non-verbal communication, socio-affective competency, interpersonal and academic skills, as well as acceleration of the children's learning rate. Apart from the 8 studies conducted in north Europe, where the therapeutic intervention was implemented primarily in school settings, only 11 of the rest of the reviewed studies provided additional information about school placement after the intervention. In 8 of those studies, 61.59% of the participants were fully included in regular school settings, with 24.39% receiving no additional support, while the rest continued to need support. Three studies reported that part of their sample had partial rather than full inclusion.

3.2. Standardized mean change effect size

The figures with the results of ESs based on standardized mean change are accompanied by forest plots (graphical representations of the results). The mean ES of each study is represented by a black square and the confidence interval (CI) with a horizontal line. The diamond at the bottom of forest-plot represents the overall ES and the diamond's width represents the overall 95% CI.

3.2.1. IQ provided by verbal standardized tests

The ES of IQ (provided by verbal test) was based on data from 21 studies. The calculation of ES_{change} showed that behavioral intervention programs were very effective (*g = 0.740, SE = 0.095, 95% CI: 0.554–0.927*) in improving children's intellectual abilities between pre-treatment and post-treatment assessments. ESs ranged from 0.190 to 1.610 (see Fig. 2).

Pertaining to heterogeneity among data for IQ provided by verbal tests, the Q-statistic showed statistically significant heterogeneity: Q(20) = 38.774, p = .007. Further analysis with I^2 statistics – depicting percentages of the total variability of ESs that could be attributed to true heterogeneity – showed moderate heterogeneity for IQ (by verbal test) (I^2 = 48.419%).

3.2.2. IQ provided by non-verbal standardized tests

The number of studies which were included for the calculation of ES_{change}, provided by non-verbal standardized tests, was 10. The effectiveness of ABA programs in improving IQ, provided by non-verbal standardized tests, is characterized moderate (*g = 0.463*, *SE = 0.129* and 95% *CI*: 0.211–0.715). ESs ranged from −0.190 to 1.310 (see Fig. 3). The heterogeneity of the ESs was not significant

Study name	Statistics for each study				
	Hedges's g	Standard error	Lower limit	Upper limit	p-Value
Anderson et al., (1987)	0.180	0.390	-0.584	0.944	0.644
Ben-Itzchak & Zachor (2007)	1.010	0.300	0.422	1.598	0.001
Bibby et al. (2002)	0.190	0.300	-0.398	0.778	0.527
Cohen et al. (2006)	1.520	0.350	0.834	2.206	0.000
Eikeseth et al. (2002)	1.100	0.420	0.277	1.923	0.009
Eldevik et al. (2006)	0.500	0.400	-0.284	1.284	0.211
Eldevik et al. (2012)	0.700	0.260	0.190	1.210	0.007
Grindle et al. (2012)	0.370	0.540	-0.688	1.428	0.493
Harris & Handleman (2000)	0.680	0.280	0.131	1.229	0.015
Harris et al. (1991)	1.090	0.510	0.090	2.090	0.033
Hayward et al. (2009)	0.890	0.220	0.459	1.321	0.000
Howard et al. (2005)	1.580	0.310	0.972	2.188	0.000
Lovaas (1987)	1.610	0.370	0.885	2.335	0.000
Peters-Scheffer et al. (2010)	0.660	0.420	-0.163	1.483	0.116
Peters-Scheffer et al. (2013)	0.370	0.320	-0.257	0.997	0.248
Remington et al. (2007)	0.530	0.300	-0.058	1.118	0.077
Rivard et al. (2014)	0.340	0.150	0.046	0.634	0.023
Sallows & Graupner (2005)	0.880	0.410	0.076	1.684	0.032
Smith et al. (1997)	0.780	0.440	-0.082	1.642	0.076
Smith et al. (2000)	0.830	0.380	0.085	1.575	0.029
Smith et al. (2015)	0.390	0.170	0.057	0.723	0.022
	0.740	0.095	0.554	0.927	0.000

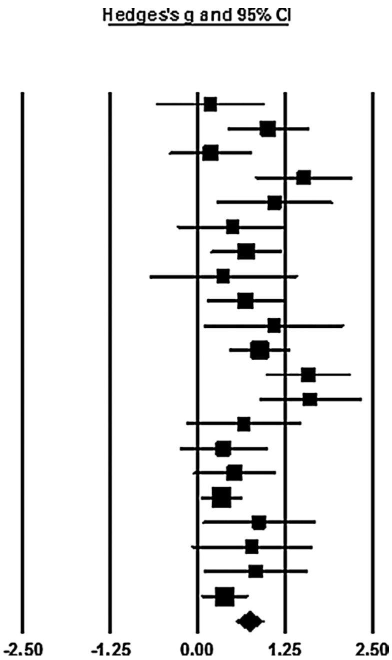

Fig. 2. Forest plot of effect sizes (Hedges's g) for IQ provided by verbal standardized tests.

Study name	Hedges's g	Standard error	Lower limit	Upper limit	p-Value
Cohen et al. (2006)	0.740	0.320	0.113	1.367	0.021
Eikeseth et al. (2002)	0.640	0.460	-0.262	1.542	0.164
Eldevik et al. (2006)	0.300	0.500	-0.680	1.280	0.549
Hayward et al. (2009)	0.420	0.220	-0.011	0.851	0.056
Howard et al. (2005)	1.310	0.330	0.663	1.957	0.000
Magiati et al. (2007)	-0.190	0.270	-0.719	0.339	0.482
Rivard et al. (2014)	0.350	0.150	0.056	0.644	0.020
Sallows & Graupner (2005)	0.320	0.390	-0.444	1.084	0.412
Sheinkopf & Siegel (1998)	1.050	0.750	-0.420	2.520	0.162
Smith et al. (2000)	0.390	0.370	-0.335	1.115	0.292
	0.463	0.129	0.211	0.715	0.000

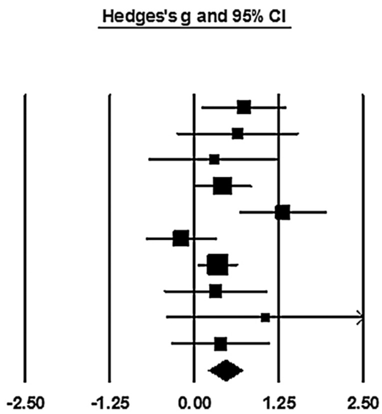

Fig. 3. Forest plot of effect sizes (Hedges's g) for IQ provided by non-verbal standardized tests.

[*Q(9) = 14.657, p = 0.101*] while I^2 statistics, which depict the percentages of the total variability of ESs that could be attributed to true heterogeneity, show low heterogeneity for IQ provided by non-verbal standardized tests (I^2 = 38.595%).

3.2.3. Adaptive behavior (composite score)

For the estimation of the standardized mean change ES of adaptive behavior in total, 17 studies were included. Findings indicated that the ABA programs were moderately effective in improving the adaptive behavior (in total) of children (*g = 0.422, SE = 0.122, 95% CI: 0.183–0.662*). ESs ranged from −0.240 to 1.590 (see Fig. 4). Data for the subscales of adaptive behavior (communication skills, socialization, daily living skills) were provided in12 studies. The estimation of relevant ESs was calculated based on those data. According to those estimations, ABA intervention programs were moderately to very effective in improving the communication skills of children with ASD (*g = 0.650, SE = 0.110, 95% CI: 0.435–0.865*, range of ESs: 0.090–1.490; see Fig. 5), moderately effective in improving socialization skills *(g = 0.444, SE = 0.085, 95% CI: 0.278–0.611*, range of ESs: *0.170–0.880*; see Fig. 6), and had low effectiveness in improving daily living skills *(g = 0.138, SE = 0.113, 95% CI: −0.083* to *0.360)* (range of ESs: −0.550 to 0.970; see Fig. 7).

Regarding the homogeneity of ESs, the Q-statistic showed statistically significant heterogeneity among data for adaptive behavior (composite scores): *Q(16) = 52.212, p < 0.001.* The variability of the ESs from the three subscales of adaptive behavior was not larger than expected on the basis of the SE [communication: *Q(11) = 16.314,* p = *0.130,* socialization: *Q(11) = 7.227, p = 0.780,* daily living skills: *Q(11) = 18.135, p = 0.079]*. Similarly, analysis with I^2 *statistics* – depicting percentages of the total variability of ESs that could be attributed to true heterogeneity – showed substantial heterogeneity for adaptive behavior (composite scores) (I^2 = 69.355%) and low heterogeneity for communication (subscale of adaptive behavior) (I^2 = 32.571%*)*, socialization (subscale of adaptive behavior) (I^2 = 0%), and daily living skills (subscale of adaptive behavior) (I^2 = 39.342%).

3.2.4. Expressive and receptive language

The ES_{change} of both expressive and receptive language is based on data from 7 studies which indicate that ABA programs are moderately to very effective in improving expressive language (*g* = 0.742*, SE* = 0.236, 95% *CI:* 0.280–1.205) and moderately

Study name	Hedges's g	Standard error	Lower limit	Upper limit	p-Value
Bibby et al. (2002)	0.480	0.310	-0.128	1.088	0.122
Cohen et al. (2006)	1.090	0.340	0.424	1.756	0.001
Eikeseth et al. (2002)	0.830	0.410	0.026	1.634	0.043
Eikeseth et al. (2012)	0.730	0.250	0.240	1.220	0.004
Eldevik et al. (2006)	-0.010	0.390	-0.774	0.754	0.980
Eldevik et al. (2012)	0.550	0.260	0.040	1.060	0.034
Femell et al. (2012)	0.020	0.150	-0.274	0.314	0.894
Flanagan et al. (2012)	0.080	0.180	-0.273	0.433	0.657
Grindle et al, 2012	1.050	0.530	0.011	2.089	0.048
Hayward et al. (2009)	0.490	0.220	0.059	0.921	0.026
Howard et al. (2005)	0.930	0.290	0.362	1.498	0.001
Magiati et al. (2007)	-0.240	0.270	-0.769	0.289	0.374
Reed et al., 2007	0.300	0.410	-0.504	1.104	0.464
Sallows & Graupner (2005)	0.450	0.400	-0.334	1.234	0.261
Smith et al. (2000)	-0.100	0.370	-0.825	0.625	0.787
Smith et al. (2015)	-0.230	0.170	-0.563	0.103	0.176
Weiss (1999)	1.590	0.360	0.884	2.296	0.000
	0.422	0.122	0.183	0.662	0.001

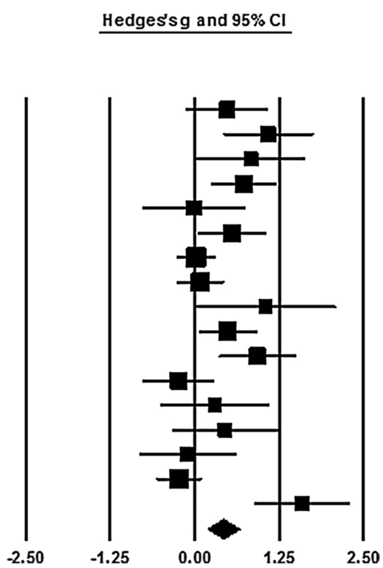

Fig. 4. Forest plot of effect sizes (Hedges's g) for composite score of adaptive behavior.

Study name	Hedges's g	Standard error	Lower limit	Upper limit	p-Value
	Statistics for each study				
Eikeseth et al. (2002)	1.140	0.420	0.317	1.963	0.007
Eikeseth et al. (2012)	0.900	0.250	0.410	1.390	0.000
Eldevik et al. (2006)	0.440	0.400	-0.344	1.224	0.271
Eldevik et al. (2012)	0.610	0.260	0.100	1.120	0.019
Grindle et al, 2012	0.950	0.530	-0.089	1.989	0.073
Hayward et al. (2009)	0.450	0.220	0.019	0.881	0.041
Howard et al. (2005)	1.490	0.310	0.882	2.098	0.000
Magiati et al. (2007)	0.090	0.270	-0.439	0.619	0.739
Reed et al., 2007	0.570	0.420	-0.253	1.393	0.175
Sallows & Graupner (2005)	0.680	0.400	-0.104	1.464	0.089
Smith et al. (2000)	0.430	0.370	-0.295	1.155	0.245
Zachor & Ben Itzchak (2010)	0.500	0.210	0.088	0.912	0.017
	0.650	0.110	0.435	0.865	0.000

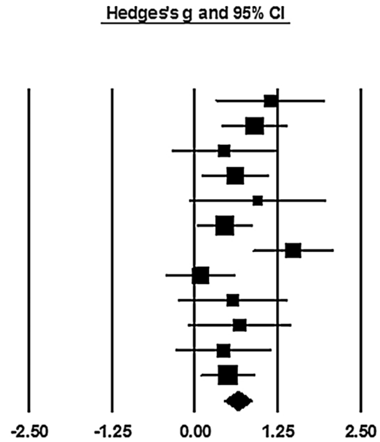

Fig. 5. Forest plot of effect sizes (Hedges's g) for communication skills (adaptive behavior).

Study name	Hedges's g	Standard error	Lower limit	Upper limit	p-Value
	Statistics for each study				
Eikeseth et al. (2002)	0.730	0.410	-0.074	1.534	0.075
Eikeseth et al. (2012)	0.630	0.240	0.160	1.100	0.009
Eldevik et al. (2006)	0.540	0.400	-0.244	1.324	0.177
Eldevik et al. (2012)	0.520	0.260	0.010	1.030	0.046
Grindle et al, 2012	0.610	0.510	-0.390	1.610	0.232
Hayward et al. (2009)	0.350	0.210	-0.062	0.762	0.096
Howard et al. (2005)	0.790	0.290	0.222	1.358	0.006
Magiati et al. (2007)	0.170	0.270	-0.359	0.699	0.529
Reed et al., 2007	0.420	0.410	-0.384	1.224	0.306
Sallows & Graupner (2005)	0.880	0.410	0.076	1.684	0.032
Smith et al. (2000)	0.210	0.370	-0.515	0.935	0.570
Zachor & Ben Itzchak (2010)	0.170	0.210	-0.242	0.582	0.418
	0.444	0.085	0.278	0.611	0.000

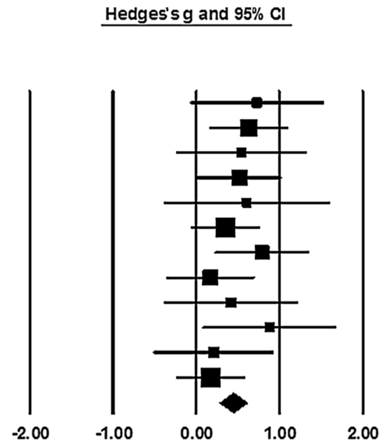

Fig. 6. Forest plot of effect sizes (Hedges's g) for socialization (adaptive behavior).

Study name	Hedges's g	Standard error	Lower limit	Upper limit	p-Value
	Statistics for each study				
Eikeseth et al. (2002)	0.660	0.400	-0.124	1.444	0.099
Eikeseth et al. (2012)	0.470	0.240	-0.000	0.940	0.050
Eldevik et al. (2006)	-0.350	0.400	-1.134	0.434	0.382
Eldevik et al. (2012)	0.170	0.250	-0.320	0.660	0.497
Grindle et al, 2012	0.970	0.530	-0.069	2.009	0.067
Hayward et al. (2009)	0.120	0.210	-0.292	0.532	0.568
Howard et al. (2005)	0.530	0.280	-0.019	1.079	0.058
Magiati et al. (2007)	-0.550	0.280	-1.099	-0.001	0.049
Reed et al., 2007	0.290	0.410	-0.514	1.094	0.479
Sallows & Graupner (2005)	0.120	0.390	-0.644	0.884	0.758
Smith et al. (2000)	-0.390	0.370	-1.115	0.335	0.292
Zachor & Ben Itzchak (2010)	0.010	0.210	-0.402	0.422	0.962
	0.138	0.113	-0.083	0.360	0.222

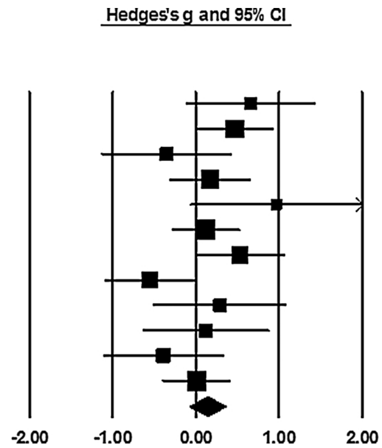

Fig. 7. Forest plot of effect sizes (Hedges's g) for daily living skills (adaptive behavior).

effective in improving receptive language (*g = 0.597, SE = 0.182, 95% CI: 0.240–0.955*) (see Figs. 8 and 9 respectively). Those results should be interpreted with caution because of the small number of studies that they were drawn from.

The Cochran's Q statistic for homogeneity showed statistically significant heterogeneity among data for expressive language: *Q (6) = 19.902, p = 0.003,* but not for receptive language: *Q(6) = 12.367, p = 0.054.* Likewise, further analysis with I^2 showed substantial heterogeneity for expressive language (I^2 = 69.852%) and moderate heterogeneity for receptive language (I^2 = 51.482%).

3.3. Publication bias

The first step in assessment for publication bias was visual inspection of the *funnel plots,* which are graphic displays of the ESs by

Study name	Hedges's g	Standard error	Lower limit	Upper limit	p-Value
Cohen et al. (2006)	1.700	0.360	0.994	2.406	0.000
Eikeseth et al. (2002)	1.380	0.480	0.439	2.321	0.004
Eldevik et al. (2006)	0.690	0.400	-0.094	1.474	0.085
Hayward et al. (2009)	0.140	0.210	-0.272	0.552	0.505
Howard et al. (2005)	0.980	0.300	0.392	1.568	0.001
Sallows & Graupner (2005)	0.230	0.390	-0.534	0.994	0.555
Smith et al. (2000)	0.310	0.370	-0.415	1.035	0.402
	0.742	0.236	0.280	1.205	0.002

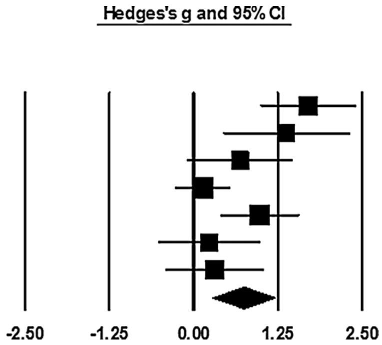

Fig. 8. Forest plot of effect sizes (Hedges's g) for expressive language.

Study name	Hedges's g	Standard error	Lower limit	Upper limit	p-Value
Cohen et al. (2006)	1.310	0.340	0.644	1.976	0.000
Eikeseth et al. (2002)	0.540	0.440	-0.322	1.402	0.220
Eldevik et al. (2006)	0.450	0.400	-0.334	1.234	0.261
Hayward et al. (2009)	0.050	0.210	-0.362	0.462	0.812
Howard et al. (2005)	0.910	0.290	0.342	1.478	0.002
Sallows & Graupner (2005)	0.630	0.400	-0.154	1.414	0.115
Smith et al. (2000)	0.440	0.370	-0.285	1.165	0.234
	0.597	0.182	0.240	0.955	0.001

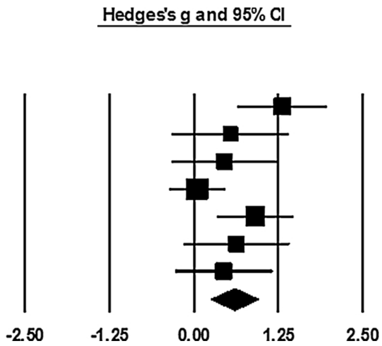

Fig. 9. Forest plot of effect sizes (Hedges's g) for receptive language.

standard errors for each variable separately (see online Supplementary Figs. 1–8). Almost all of them were symmetric and consequently, there was no significant evidence for publication bias. In addition, the nonparametric method of Begg and Mazumdar did not yield statistically significant publication bias for any one of the variables (IQ provided by verbal test *Kendall's tau = 0.138, p = 0.191*, IQ provided by non-verbal test *Kendall's tau = 0.178, p = 0.237,* adaptive behavior *Kendall's tau = 0.279, p = 0.059,* communication-subscale of adaptive behavior *Kendall's tau = 0.288, p = 0.096,* socialization-subscale of adaptive behavior *Kendall's tau = 0.258, p = 0.122,* daily living skills-subscale of adaptive behavior *Kendall's tau = 0.182, p = 0.205,* expressive language *Kendall's tau = 0.190, p = 0.274, and* receptive language *Kendall's tau = 0.048, p = 0.440*).

Subsequent statistical analyses, using *Egger's* linear *regression,* indicated possible bias for two of the variables. Specifically, the Egger's intercept for IQ provided by verbal tests (Egger's intercept *BO = 1.680, p = 0.027*) and adaptive behavior (Egger's intercept *BO = 2.890, p = 0.009) were significant, indicating potential publication bias*. For the rest of the variables, there was no significant indication of publication bias. Specifically, no bias was identified for IQ (provided by non-verbal test) (Egger's intercept *BO = 0.908, p = 0.385*), communication-subscale of adaptive behavior (*BO = 1.182, p = 0.385*), socialization-subscale of adaptive behavior (*BO = 1.224, p = 0.158*), daily living skills-subscale of adaptive behavior (*BO = 0.563, p = 0.697*), expressive language (*BO = 3.535, p = 0.201*), and for receptive language (*BO = 2.487, p = 0.214*) (funnel plots are provided on line).

4. Discussion and implications

The aim of the present meta-analysis was to estimate the effectiveness of ABA programs for children with ASD. Analysis of ESs suggested that ABA programs were very effective in improving the intellectual abilities; moderately to very effective in improving the communication skills and expressive and receptive language skills; moderately effective in improving IQ scores yielded by non-verbal tests, adaptive behavior (in total), socialization, and receptive language skills. Lastly, the smaller gains were obtained in the domain of daily living skills. Those smaller gains may be attributed to the young age of the participants of the meta-analyzed studies as well as the emphasis of ABA programs on language, communication, and cognitive skills more so than other developmental areas. Nevertheless, as children with ASD grow older, the emphasis on daily living skills should be greater as it may enhance their ability to be self-sufficient and independent of continuous support from their caregivers.

In general, the findings of the present meta-analysis support the effectiveness of ABA intervention similarly to previous meta-analytic studies and reviews, with the exception of daily living skills for which there was high heterogeneity in outcomes among the meta-analytic studies. Despite the common general conclusion of all meta-analytic studies about effectiveness, a direct comparison of their findings is not easily attainable for several reasons: Firstly, because of great variation in research procedures and designs used by different meta-analytic studies. For example, in some, the effectiveness of ABA intervention comparing the gains of the experimental with a comparison group were studied (Eldevik et al., 2009; Peters-Scheffer et al., 2011; Reichow et al., 2012; Spreckley & Boyd, 2009), in another study pre- and post-treatment performance was assessed (i.e., Virues-Ortega, 2010) and yet in two others both types – experimental-control performance and pre-post-treatment performance – were assessed (Makrygianni & Reed, 2010; Reichow, &

Wolery, 2009).

A second reason that prevents a possible comparison of the findings of different meta-analyses is the use of different measures of ES, such as Hedges' *g*, Cohen's *d* and others, which yield different ES values. Thirdly, meta-analytic studies usually adopt different inclusion criteria which may lead to including different sets of studies. For example, among the meta-analytic studies reviewed for the purposes of this study, the sample of included studies varied from 4 (Spreckley & Boyd, 2009) to 22 (Virues-Ortega, 2010). With such variability, it is not possible to compare the ES values of different meta-analytic studies. As the number of meta-analyzed studies included in a review increases, so does the possibility of including studies that do not have sound methodologies. Thus, including a large number of studies may bias the outcome of a meta-analysis and specifically the calculation of ESs.

Lastly, there are various methodological flaws that may attribute to inaccurate meta-analytic outcomes, such as using age-equivalent or raw scores (e.g., Reichow et al., 2012) instead of using standard scores to calculate ESs, or not providing sufficient description of the methodologies used which prevents an assessment of the accuracy or validity of the meta-analytic results (e.g., Peters-Scheffer et al., 2011; Reichow & Wolery, 2009).

Despite the practical difficulties associated with comparing meta-analytic studies, it is important to underline that all of them point out to the effectiveness of ABA interventions for the treatment of ASD. Yet, there is variability among studies pertaining to the extent of treatment effectiveness as well as to gains obtained across different developmental domains. When interpreting meta-analytic results, all of the aforementioned factors should be taken into account.

Limitations of the present study are similar to those usually associated with comprehensive syntheses of studies. Firstly, those associated with methodological aspects of the included studies, such as small sample size, not using a random assignment of children to groups, poorly designed ABA programs, or lack of evidence that the investigators controlled threats to the internal validity (e.g., maturation factors that may have attributed to treatment gains). A second limitation has to do with the variability among the included studies. Meta-analytic research calls for the maximum possible homogeneity of the included studies. Yet, the studies included in the present meta-analysis varied across several factors, such as age of the participants, participants' characteristics at intake, ABA therapeutic protocols used, intensity of intervention, and parental involvement in the treatment. Another limitation that was difficult to avoid has to do with the lack of adequate number of studies that included certain variables related to child performance, such as expressive and receptive language and IQ scores. Even though the present meta-analysis incorporated a greater number of studies that included these variables, compared to those used in prior meta-analyses for the calculation of ESs, the number of such studies continues to be very small. Finally, the present meta-analysis did not include a post comparison between control and experimental groups which is common practice in relevant meta-analytic studies.

For future meta-analytic research, several issues may be taken into consideration. First of all, when the evaluation of the effectiveness of the intervention is based on comparisons of Standard Scores (SSs) before and after ABA intervention, we may underestimate the size of the effectiveness of the intervention for pupils with ASD whose learning rates are below average. For example, in the Bibby, Eikeseth, Martin, Mudford, and Reeves (2002) study, participants with autism receiving a parent-managed behavioral intervention program for 12 months demonstrated progress, in daily living skills, equivalent to 7.6 chronological months. Yet, based on comparisons of SSs this improvement was not depicted. Rather, the change of SSs indicated a 2.1-point decrease in the area of daily living skills. The same is true for the Magiati and her colleagues' study (Magiati, Charman, & Howlin, 2007). Although they reported reduction of adaptive behavior after 25 months of intervention, it is worth noting that the participants' age equivalent actually increased during that time by 18 chronological months. So, we may consider age-equivalent scores as a more accurate and appropriate way to evaluate intervention effectiveness for children with ASD than standard or ratio scores.

In conclusion, the present meta-analysis demonstrates the degree of effectiveness of ABA programs across different developmental domains. The findings of this meta-analysis provide further support for the effectiveness of the use of ABA methodologies for the treatment of children with ASD. Yet, we are not aware of the components of ABA programs that may be most effective in producing optimal therapeutic outcomes (e.g., improvement in IQ scores, school inclusion, etc., Fein et al., 2013). Such an assessment is not easily attained unless intervention studies are meticulously designed and executed (e.g., use: random rather than convenient samples, a randomized control trial methodology, blind assessment, the same instruments to evaluate the dependent measures before and after treatment, etc.). Yet, methodological rigor of intervention studies may not always be achieved especially for populations of children with compelling needs for intervention, as is the case for children with ASD.

Conflict of interest statement

All authors certify that they have no conflict of interest in this study to disclose.

Acknowledgments

This study was funded by the National and Kapodistrian University of Athens (www.uoa.gr), Greece – 2012-13 Department of Philosophy, Education, and Psychology.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.rasd.2018.03.006.

References[1]

*Anderson, S. R., Avery, D. L., DiPietro, E. K., Edwards, G. L., & Christian, W. P. (1987). Intensive home-based early intervention with autistic children. *Education and Treatment of Children, 10*, 352–366.
Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., Meldrum, D., & Charman, T. (2006). Prevalence of the disorders of the autism spectrum in a population cohort of children in South Thames: The Special Needs and Autism Project (SNAP). *The Lancet, 368*(9531), 210–215.
Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*, 1088–1101.
*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 outcome for children with autism receiving parent- managed intensive interventions. *Research in Developmental Disabilities, 23*, 81–104.
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. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*, 97–111. http://dx.doi.org/10.1002/jrsm.12.
*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.
Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple graphical test. *British Medical Journal, 315*, 629–634.
*Eikeseth, S., Smith, T., Jahr, E., & Eldevik, S. (2002). Intensive behavioural 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., Klintwall, L., Jahr, E., & Karlsson, P. (2012). Outcome for children with autism receiving early and intensive behavioral intervention in mainstream preschool and kindergarten settings. *Research in Autism Spectrum Disorders, 6*, 829–835. http://dx.doi.org/10.1016/j.rasd.2011.09.002.
*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 and Developmental Disorders, 36*, 211–224.
Eldevik, S., Hastings, R. P., Hughes, 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 participant data to extend the evidence base for intensive behavioral intervention for children with autism. *American Journal on Intellectual and Developmental Disabilities, 115*(5), 381–405.
*Eldevik, S., Hastings, R. P., Jahr, E., & Hughes, J. C. (2012). Outcomes of behavioral intervention for children with autism in mainstream pre-school settings. *Journal of Autism and Developmental Disorders, 42*, 210–220. http://dx.doi.org/10.1007/s10803-011-1234-9.
Fava, L., Strauss, K., Valeri, G., D'Elia, L., Arima, S., & Vicari, S. (2011). The effectiveness of a cross-setting complementary staff-and parent-mediated early intensive behavioral intervention for young children with ASD. *Research in Autism Spectrum Disorders, 5*, 1479–1492. http://dx.doi.org/10.1016/j.rasd.2011.02.009.
Fein, D., Barton, M., Eigsti, I. M., Kelley, E., Naigles, L., Schultz, R. T., ... Tyson, K. (2013). Optimal outcome in individuals with a history of autism. *Journal of Child Psychology and Psychiatry, 54*, 195–205. http://dx.doi.org/10.1111/jcpp.12037.
*Fernell, E., Hedvall, A., Westerlund, J., Carlsson, L. H., Eriksson, M., Olsson, M. B., ... Gillberg, C. (2011). Early intervention in 208 Swedish preschoolers with autism spectrum disorder. A prospective naturalistic study. *Research in Developmental Disabilities, 32*, 2092–2101. http://dx.doi.org/10.1016/j.ridd.2011.08.002.
*Flanagan, H. E., Perry, A., & Freeman, N. L. (2012). Effectiveness of large-scale community-based intensive behavioral Intervention: A waitlist comparison study exploring outcomes and predictors. *Research in Autism Spectrum Disorders, 6*, 673–682. http://dx.doi.org/10.1016/j.rasd.2011.09.011.
*Grindle, C. F., Hastings, R. P., Saville, M., Hughes, J. C., Huxley, K., Kovshoff, H., ... Remington, B. (2012). Outcomes of a behavioral education model for children with autism in a mainstream school setting. *Behavior Modification, 36*, 298–319. http://dx.doi.org/10.1177/0145445512441199.
*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 and 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 and Developmental Disorders, 21*, 281–290.
*Hayward, D., Eikeseth, S., Gale, C., & Morgan, S. (2009). Assessing progress during treatment for young children with autism receiving intensive behavioural interventions. *Autism, 13*, 613–633.
Healy, O., & Lydon, S. (2013). Early intensive behavioural intervention in autism spectrum disorders. In M. Fitzgerald (Vol. Ed.), *Recent advances in autism spectrum Disorder: Vol. 1*, (pp. 567–597). INTECH Open Access Publisher. [Retrieved from] https://www.intechopen.com/books/recent-advances-in-autism-spectrum-disorders-volume-i.
Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.
Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics, 6*, 107–128.
Cochrane handbook for systematic reviews of interventions. In J. P. Higgins, & S. Green (Vol. Eds.), *Vol. 4*. Chichester: John Wiley & Sons.
Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*, 1539–1558.
*Howard, J. S., Sparkman, C. R., Cohen, H. G., Green, G., & Stanislaw, H. (2005). A comparison of intensive behaviour analytic and eclectic treatments for young children with autism. *Research in Developmental Disabilities, 26*, 359–383.
Howlin, P., Magiati, I., & Charman, T. (2009). Systematic review of early intensive behavioral interventions for children with autism. *American Journal on Intellectual and Developmental Disabilities, 114*.
Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods, 11*, 193–206.
Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods, 17*, 137–152. http://dx.doi.org/10.1037/a0028086.
Kim, Y. S., Leventhal, B. L., Koh, Y. J., Fombonne, E., Laska, E., Lim, E. C., ... Song, D. H. (2011). Prevalence of autism spectrum disorders in a total population sample. *American Journal of Psychiatry, 168*, 904–912. http://dx.doi.org/10.1176/appi.ajp.2011.10101532.
Kuppens, S., & Onghena, P. (2012). Sequential meta-analysis to determine the sufficiency of cumulative knowledge: The case of early intensive behavioral intervention for children with autism spectrum disorders. *Research in Autism Spectrum Disorders, 6*, 168–176. http://dx.doi.org/10.1016/j.rasd.2011.04.002.
Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Los Angeles: Sage Publications.
*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.
*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.
Makrygianni, M. K., & Reed, P. (2010). A meta-analytic review of the effectiveness of behavioural early intervention programs for children with autism spectrum disorders. *Research in Autism Spectrum Disorders, 4*, 577–593. http://dx.doi.org/10.1016/j.rasd.2010.01.014.
McEachin, J. J., Smith, T., & Lovaas, O. I. (1993). Long-term outcome for children with autism who received early intensive behavioral treatment. *American Journal of Mental Retardation, 97*, 359–372.
O'Mara, A. J., Marsh, H. W., & Craven, R. G. (2005). Meta-Analytic methods in educational research: Issues and their solutions. *Paper presented at Australian Association for research in education annual conference*. [Retrieved from] http://www.aare.edu.au/data/publications/2005/oma05407.pdf.
*Peters-Scheffer, N., Didden, R., Mulders, M., & Korzilius, H. (2010). Low intensity behavioral treatment supplementing preschool services for young children with

[1] Marked with an asterisk (*) indicate studies included in present meta-analysis.

autism spectrum disorders and severe to mild intellectual disability. *Research in Developmental Disabilities, 31*, 1678–1684. http://dx.doi.org/10.1016/j.ridd.2010.04.008.
Peters-Scheffer, N., Didden, R., Korzilius, H., & Sturmey, P. (2011). A meta-analytic study on the effectiveness of comprehensive ABA-based early intervention programs for children with autism spectrum disorders. *Research in Autism Spectrum Disorders, 5*, 60–69. http://dx.doi.org/10.1016/j.rasd.2010.03.011.
*Peters-Scheffer, N., Didden, R., Mulders, M., & Korzilius, H. (2013). Effectiveness of low intensity behavioral treatment for children with autism spectrum disorder and intellectual disability. *Research in Autism Spectrum Disorders, 7*, 1012–1025. http://dx.doi.org/10.1016/j.rasd.2013.05.001.
*Reed, P., Osborne, L. A., & Corness, M. (2007). The real-world effectiveness of early teaching interventions for children with autism spectrum disorder. *Exceptional Children, 73*, 417–433.
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.
Reichow, B., Barton, E. E., Boyd, B. A., & Hume, K. (2012). Early intensive behavioral intervention (EIBI) for young children with autism spectrum disorders (ASD). *Cochrane Database of Systematic Reviews,* 10. http://dx.doi.org/10.1002/14651858.CD009260. pub2.
Reichow, B. (2012). Overview of meta-analyses on early intensive behavioral intervention for young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 42*, 512–520. http://dx.doi.org/10.1007/s10803-011-1218-9.
*Remington, B., Hastings, R. P., Kovshoff, H., degliEspinosa, 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.
*Rivard, M., Terroux, A., & Mercier, C. (2014). Effectiveness of early behavioral intervention in public and mainstream settings: The case of preschool-age children with autism spectrum disorders. *Research in Autism Spectrum Disorders, 8*, 1031–1043.
Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118*, 183–192.
Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2006). *Publication bias in meta-analysis: Prevention, assessment and adjustments.* New Jersey: Wiley & Sons.
Saemundsen, E., Magnússon, P., Georgsdóttir, I., Egilsson, E., & Rafnsson, V. (2013). Prevalence of autism spectrum disorders in an Icelandic birth cohort. *BMJ Open, 3*(6), e002748. http://dx.doi.org/10.1136/bmjopen-2013-002748.
*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.
*Sheinkopf, S. J., & Siegal, B. (1998). Home- based behavioural treatment of young children with autism. *Journal of Autism and Developmental Disorders, 28*, 15–23.
*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 of 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.
*Smith, T., Klorman, R., & Mruzek, D. W. (2015). Predicting outcome of community-based early intensive behavioral intervention for children with autism. *Journal of Abnormal Child Psychology, 43*, 1271–1282.
Spreckley, M., & Boyd, R. (2009). Efficacy of applied behavioral intervention in preschool children with autism for improving cognitive, language, and adaptive behavior: A systematic review and meta-analysis. *The Journal of Pediatrics, 154*, 338–344.
Sterne, J. A., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology, 54*, 1046–1055.
Strauss, K., Vicari, S., Valeri, G., D'Elia, L., Arima, S., & Fava, L. (2012). Parent inclusion in early intensive behavioral intervention: The influence of parental stress, parent treatment fidelity and parent-mediated generalization of behavior targets on child outcomes. *Research in Developmental Disabilities, 33*, 688–703. http://dx.doi.org/10.1016/j.ridd.2011.11.008.
Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods, 1*, 112–125. http://dx.doi.org/10.1002/jrsm.11.
Virues-Ortega, J. (2010). Applied behavior analytic intervention for autism in early childhood: Meta-analysis, meta-regression and dose-response meta-analysis of multi- ple outcomes. *Clinical Psychology Review, 30*, 387–399. http://dx.doi.org/10.1016/j.cpr.2010.01.008.
Warren, Z., McPheeters, M. L., Sathe, N., Foss-Feig, J. H., Glasser, A., & Veenstra-VanderWeele, J. (2011). A systematic review of early intensive intervention for autism spectrum disorders. *Pediatrics, 127*, 1303–1311. http://dx.doi.org/10.1542/peds.2011-0426.
*Weiss, M. J. (1999). Differential rates of skill acquisition and outcomes of early intensive behavioral intervention for autism. *Behavioral Interventions, 14*, 3–22.
*Zachor, D. A., & Ben-Itzchak, E. (2010). Treatment approach, autism severity and intervention outcomes in young children. *Research in Autism Spectrum Disorders, 4*(3), 425–432. http://dx.doi.org/10.1016/j.rasd.2009.10.013.