Clinical Psychology Review 30 (2010) 387–399

Contents lists available at ScienceDirect

Clinical Psychology Review

Applied behavior analytic intervention for autism in early childhood: Meta-analysis, meta-regression and dose–response meta-analysis of multiple outcomes

Javier Virués-Ortega *

Instituto de Salud Carlos III, CIBERNED, Madrid, Spain
ABA España, Madrid, Spain

ARTICLE INFO

Article history:
Received 13 May 2009
Received in revised form 30 December 2009
Accepted 29 January 2010

Keywords:
Autism spectrum disorders
Applied behavior analysis
Language
Meta-analysis

ABSTRACT

A number of clinical trials and single-subject studies have been published measuring the effectiveness of long-term, comprehensive applied behavior analytic (ABA) intervention for young children with autism. However, the overall appreciation of this literature through standardized measures has been hampered by the varying methods, designs, treatment features and quality standards of published studies. In an attempt to fill this gap in the literature, state-of-the-art meta-analytical methods were implemented, including quality assessment, sensitivity analysis, meta-regression, dose–response meta-analysis and meta-analysis of studies of different metrics. Results suggested that long-term, comprehensive ABA intervention leads to (positive) medium to large effects in terms of intellectual functioning, language development, acquisition of daily living skills and social functioning in children with autism. Although favorable effects were apparent across all outcomes, language-related outcomes (IQ, receptive and expressive language, communication) were superior to non-verbal IQ, social functioning and daily living skills, with effect sizes approaching 1.5 for receptive and expressive language and communication skills. Dose-dependant effect sizes were apparent by levels of total treatment hours for language and adaptation composite scores. Methodological issues relating ABA clinical trials for autism are discussed.

Contents

1. Introduction

Applied behavior analysis is a behavioral science devoted to the experimental study of socially significant behavior as a function of environmental variables. Throughout the last four decades a number of procedures aimed at enhancing, reducing and maintaining significant human behaviors have been developed by applied behavior analysts (Cooper, Heron, & Heward, 2007a). This research has had a significant impact in the fields of severe problem behavior, developmental disabilities, organizational behavior, behavioral pharmacology, behavioral economics and others. The field of applied behavior analysis has shown a more significant growth in the area of behavioral intervention

* Centro Nacional de Epidemiología, Instituto de Salud Carlos III, Sinesio Delgado 6, 28029 Madrid, Spain. Tel.: +34 134 91 822 2651; fax: +34 134 91 387 7815.
E-mail address: jvortega@cienciaconducta.com.

0272-7358/$ – see front matter
doi:10.1016/j.cpr.2010.01.008

for children with autism and autism spectrum disorders as suggested by the increasing number of service providers and certified professionals in this field (Cooper, Heron, & Heward, 2007b; Shook, 2005). Since the mid-80s (Fenske, Zalenski, Krantz, & McClannahan, 1985; Lovaas, 1987) the evidence suggesting that applied behavior-analytic intervention (hereafter referred to as ABA intervention) is beneficial to the intellectual, verbal, and social functioning of children with autism and autism spectrum disorders has accumulated steadily (Foxx, 2008; Remington et al., 2007).

Although there are several models of ABA intervention in autism and developmental disabilities, all bonafide programs should share a common set of core features: (1) treatment may begin as early as 3 to 4 years of age, (2) intervention is intensive (20–40 weekly hours) and in addition, incidental teaching and practice goals may be operating during most waking hours, (3) intervention is individualized and comprehensive targeting a wide range of skills, (4) multiple behavior analytic procedures are used to develop adaptive repertoires, (5) treatment is delivered in one-to-one format with gradual transition to group activities and natural contexts, (6) treatment goals are guided by normal developmental sequences, and (8) parents are, to different extents, trained and become active co-therapists (Maurice, Green, & Foxx, 2001).

Positive results have been reported for daily living skills, academic performance and communication skills (Eikeseth, Smith, Jahr, & Eldevik, 2007; Remington et al., 2007). Studies suggest that with ABA intervention, children have a greater chance of integrating into school without additional specialist support whilst maintaining gains over long follow-up periods (McEachin, Smith, & Lovaas, 1993). These findings have had some effects on the social and health policies of different countries (New York State Department of Health, 1999; Ontario Ministry of Education, 2007; U.S. Public Health Service, 1999). However, dissemination of research findings may still be considered limited. For example, recent reviews on autism do not even acknowledge the very existence of ABA intervention (Hughes, 2008) or misrepresent its application and effects (Volkmar & Davies, 2003).

Although a number of studies have been conducted to explore the effectiveness of ABA intervention in children with developmental disabilities, the collective examination of this literature is hampered by a number of factors: (1) studies implement inconsistent methodological features with regard to research design, sampling methods and quality standards, (2) intervention features are highly variable including treatment intensity, duration, the intervention model itself and format of treatment delivery (e.g. clinic-based vs. parent-managed), (3) participants are highly variable with regard to their pre-intervention functioning and age and, finally (4) studies use of variety of different metrics when reporting outcomes making it difficult to implement standard meta-analytical procedures (Morris & DeShon, 2002). Furthermore, the fact that most literature in this area has been single-subject design research and that studies are often procedure-specific (in terms of approaches to treatment) has prevented wider dissemination of results through standard methods of clinical science. Although attempts have been made to summarize single-subject research, these methods are still controversial (Scruggs & Mastropieri, 1998; Severtson, Carr, & Lepper, 2009).

A precise quantification of ABA intervention effectiveness is not currently available. Previous reviews have focused on very specific aspects of ABA intervention (Delprato, 2001), or have failed to incorporate advanced meta-analytical procedures including quality assessment, meta-regression, dose–response meta-analysis, and meta-analysis of studies of different metrics (Eldevik, Hastings, Hughes, Jahr, & Eikeseth, 2009). The present study has the following goals: (1) ascertain the collective effectiveness of ABA intervention for autism, (2) estimate ABA intervention effectiveness in terms of as many outcome variables as possible in order to provide a comprehensive assessment of its effects, and (3), analyze the effect of study characteristics including intervention duration and intensity, study design, intervention model and intervention delivery format. This study pursues a comprehensive account of the effects of comprehensive, intensive and long-term ABA intervention over subjects' functioning in molar skills domains, therefore, studies targeting specific behaviors or procedures will be discarded.

2. Methods

2.1. Literature search and study selection

MEDLINE, PsycINFO, and the Cochrane Clinical Trials databases were searched for all studies reporting the effect of intensive, long-term ABA intervention with children with autism and pervasive developmental disabilities not otherwise specified. Although ABA intervention focuses on specific skills and behaviors at a time, as we examined the molar effects of long-term, comprehensive ABA intervention, no specific behavior or behavior procedure could be contemplated as an inclusion criterion in the assumption that they were many throughout the treatment process. Formal search strategies for randomized controlled trials were supplemented with less restrictive search strategies in order to enhance the detection of low impact journals and mid-to-low quality studies (Botella & Gambara, 2006; Robinson & Dickersin, 2002) (see search strategy in Appendix A). The search period was January 1985 through April 2009, with no language restrictions. The reference lists of selected review articles were also reviewed (British Columbia Office of Health Technology Assessment, 2001; Eldevik et al., 2009).

A number of pre-specified exclusion criteria were used to identify key studies. The 11 exclusion criteria were: (1) the study was non peer-reviewed, non-original, non-empirical, methodological or unpublished; (2) none of the intervention groups implemented ABA intervention for autism according to major features of comprehensive behavior-analytic intervention for autism (Maurice et al., 2001); (3) the focus of the intervention was for very specific areas (e.g., joint attention, problem behavior) or was restricted to a specific behavioral procedure (e.g., functional communication treatment, non-contingent reinforcement); (4) intervention did not meet the intensity and duration standards of ABA interventions (at least 10 weekly hours and no less than 45 weeks duration); (5) participants did not have a formal diagnosis of autism according to the Autism Diagnostic Interview-Revised (Lord, Rutter, & Le Couteur, 1994), the Autism Diagnostic Observation Schedule (Lord, Rutter, DiLavore, & Risi, 1999), the Diagnostic and Statistical Manual of Mental Disorders criteria for autism (American Psychiatric Association, 2000) or a combination of any of these methods; (6) the study utilized a single-subject study design or had an intervention group with less than five subjects; (7) the study was epidemiological; (8) the study reported anecdotal, qualitative or non-standardized outcome measures; (9) there was no pre-test measurement; (10) the study purposely biased subject selection (e.g., fast learners), and (11) mean and standard deviations were not available after attempts to contact authors and could not be calculated from descriptive data or statistical tests in the study manuscript. Exclusion criteria were implemented successively. Although a minimum of 10 weekly hours may be considered too low, this criterion may enable the more precise determination of the impact of intervention intensity on treatment effectiveness. Outcomes reported in less than three clinical trials were discarded from the meta-analysis. The selection process is summarized in the flow chart in Fig. 1.

2.2. Assessment of studies and data extraction

Two investigators (JV-O, MR-M) independently screened the titles and abstracts of the database searches and retrieved articles to determine eligibility (by virtue of the exclusion criteria) before extracting study data. Interrater agreement in the final number of

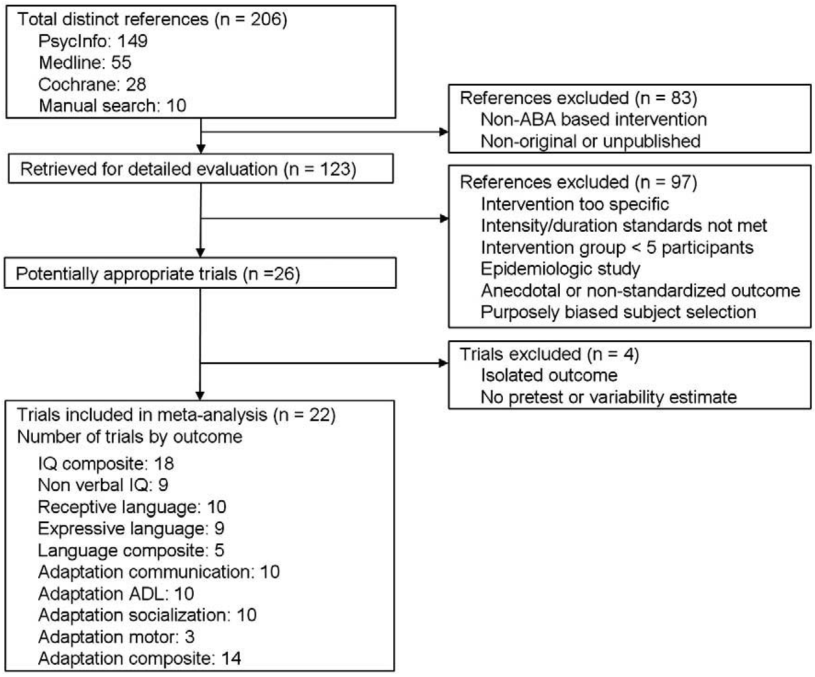

Fig. 1. Flow chart of trial selection process.

trials to be included in the meta-analysis reached 90.9%. Discrepancies were resolved by consensus. The authors of the original studies were contacted if relevant data were not available in the published reports. The service of an assistant translator, for studies published in languages other than English and Spanish, was used when necessary.

The following data were retrieved from all selected studies: (1) participant characteristic including mean pre-intervention age in months, percentage of male participants, pre-intervention IQ, (2) intervention characteristics including intervention intensity (weekly hours); intervention duration (weeks); total intervention duration (intensity multiplied by duration); intervention delivery format, whether clinic-based or parent-managed programs delivered at home and supervised by professionals (i.e., clinic-based vs. parent-managed programs), model of ABA intervention (UCLA model [Lovaas, 1981] vs. general applied behavior analytic model [e.g., Cooper et al., 2007a; Maurice, Green, & Luce, 1996]); study design (randomized controlled trial, non-randomized controlled trial, repeated measures study); sample size; outcome variables; assessment instruments; reported pre- and post-test outcome values (mean and standard deviation); and study quality. Two trained investigators (JV-O, MR-M) assessed the quality of the studies independently by means of the Downs and Black checklist for randomized and non-randomized studies of health care interventions (Downs & Black, 1998). Cohen's kappa for studies' total score reached 0.95. Discrepancies were resolved by consensus. Quality domains covered by the checklist are: Reporting, External validity, Internal validity-bias, Internal validity-confounding and Power. Domains were rated on a 0-1 scale in order to provide a 5-point total quality range and to avoid over-representation of scale domains holding more items (e.g., Reporting). As suggested by the original authors, the checklist was adapted specifically for the search topic by adding a list of confounders, adverse effects, and ranges for power assessment (see the scale and quality assessment in Appendix A). The quality checklist was selected because it was flexible enough to be applicable to both repeated measures and control group studies, whether randomized or not. Assessors' disagreements on these quality measures were resolved by consensus.

2.3. Statistical analysis

Because the instruments for evaluating a given outcome differed across studies (e.g., Wechsler Intelligence Scale for Children vs. Merrill-Palmer Scales of Mental Tests), we used effect sizes to obtain standardized measurements of the effect of the intervention on the outcome variable. For studies with a control group, effect sizes were calculated as the difference in outcome progression (that is, post- minus pre-test mean scores) between the intervention and control groups, divided by the pre-test standard deviation pooled across groups. For these studies, the intervention group comprised all participants receiving ABA intervention and the control group comprised all participants not receiving ABA intervention, irrespective of the concurrent use of other treatments and the alternative intervention assigned to the control group. For within-subjects designs, effect sizes were computed by dividing the mean difference between post- and pre-test outcomes by the pre-test standard deviation. Assuming that outcome changes at follow-up are the effect of treatment, effect size estimates from within-subjects studies are equivalent and comparable to those from controlled studies (Morris & DeShon, 2002), and they can be interpreted as the effect of the intervention on the outcome measured in pre-test within-group standard deviation units. Sensitivity analyses restricting meta-analysis to controlled studies were conducted in order to test this assumption. Once effect sizes were obtained from means and standard deviations results were combined across studies. The above effect size estimates were corrected for small-sample bias, and design-specific estimates of their sampling variance were computed (Becker, 1988; Morris, 2008). If not explicitly reported, outcome means and standard deviations were calculated from the available descriptive data or test statistics using standard methods (Morris & DeShon, 2002). Since the correlation between pre- and post-test outcomes is required to compute the effect size variance, a pooled correlation coefficient was estimated from studies in which sufficient data were available to calculate pre-post correlation coefficients for a given outcome (Morris & DeShon, 2002). The pooled estimate was then applied to all studies reporting the outcome. Interim measures were always discarded, selecting the pre-test and post-test measures closest to the beginning

Table 1
Studies reporting the effects of comprehensive and intensive applied behavior analytic intervention for autism.

First author, year	Country	Diagnosis	Male (%)	Mean age (months)	Pre-IQ	Control group	Sample size*	ABA Intervention Model	Intensity (h/week)	Duration (weeks)	Outcome (instrument)	Quality score†
Clinic-based intervention programs												
Ben-Itzchak and Zachor (2007)	Israel	Autism	92.00	26.60	70.67	No	25	General	35.00	53.00	IQ composite (BSID-II, SBIS) Receptive (BO) Expressive (BO)	1.77
Ben-Itzchak et al. (2008)	Israel	Autism	97.67	27.29	74.84	Yes	39/37	General	45.00	53.00	IQ composite (BSID-II, SBIS)	3.04
Birnbrauer and Leach (1993)	Australia	Autism	55.56	39.00	51.28	Yes ‡	9/5	UCLA	18.72	105.12	IQ composite (BSID-II, SBIS, LEITER, PPVT) Language composite (RDLS, REEL)	2.30
Cohen et al. (2006)	United States	Autism, PDD NOS	83.33	31.70	60.50	Yes, R	21 /21	UCLA	37.50	141.00	IQ composite (BSID-II, WPPSI-R, WISC-III) Non-verbal IQ (MPSMT) Receptive (RDLS) Expressive (RDLS) Adaptation-C (VABS) Adaptation-DLS (VABS) Adaptation-S (VABS) Adaptation composite (VABS)	2.90
Eikeseth et al. (2002, 2007)	Norway	Autism	80.00	66.31	65.68	Yes, R	13/12	UCLA	23.50	148.10	IQ composite (BSID-II, WPPSI-R, WISC-III) Non-verbal IQ (MPSMT) Receptive (RDLS) Expressive (RDLS) Language composite (RDLS) Adaptation-C (VABS) Adaptation-DLS (VABS) Adaptation-S (VABS) Adaptation composite (VABS)	2.97
Eldevik et al. (2006)	Norway	Autism	85.71	50.86	44.32	Yes	13/15	General	12.00	88.91	IQ composite (BSID-II, SBIS, WPPSI-R, WISC-III) Non-verbal IQ (MPSMT) Receptive (RDLS) Expressive (RDLS) Adaptation-C (VABS) Adaptation-DLS(VABS) Adaptation-S (VABS) Adaptation composite (VABS)	2.26
Harris et al. (1991)	United States	Autism	88.24	47.40	65.56	Yes	16/12	UCLA	–	49.14	IQ composite (SBIS-IV) Language composite (PLS)	2.03
Harris and Handleman (2000)	United States	Autism	85.19	49.00	59.33	No	27	General	40.00	407.34	IQ composite (SBIS-IV)	3.25
Howard et al. (2005)	United States	Autism, PDD NOS	84.44	33.20	56.69	Yes	26/16	General	32.50	62.24	IQ composite (BSID-II, WPPSI, WISC, DP-II, SBIS) Non-verbal IQ (MPSMT) Receptive (RDLS) Expressive (RDLS) Adaptation-C (VABS) Adaptation-DLS (VABS) Adaptation-S (VABS) Adaptation-motor (VABS) Adaptation composite (VABS)	2.86
Lovaas (1987)	United States	Autism	84.21	34.60	54.34	No ‡	19	UCLA	40.00	106.00	IQ composite (WPPSI, WISC, SBIS, CIIS, BSID-II, MPSMT, LEITER)	2.38
Magiati et al. (2007)	United Kingdom	Autism	88.64	39.64	76.53	Yes	28/16	UCLA	32.40	109.50	Non-verbal IQ (MPSMT) Receptive (BPLS-II) Expressive (BPLS-II) Adaptation-C (VABS) Adaptation-DLS (VABS) Adaptation-S (VABS) Adaptation composite (VABS)	2.80

Matos and Mustaca, 2005	Argentina	Autism, PDD NOS	88.89	42.00	15.00	No ‡	9	General	40.00	48.18	IQ composite (BSID-II) Receptive (PPVT) Adaptation-C (VABS) Adaptation-DLS (VABS) Adaptation-S (VABS) Adaptation-motor (VABS) Adaptation composite (VABS)	1.17
Remington et al. (2007)	United Kingdom	Autism	–	37.10	61.90	Yes	23/21	General	25.60	105.12	IQ composite (BSID-II, SBIS) Adaptation-C (VABS) Adaptation-DLS (VABS) Adaptation-S (VABS) Adaptation-motor (VABS) Adaptation composite (VABS)	2.67
Sallows and Graupner (2005)	United States	Autism	84.62	33.23	50.85	No ‡	13/10	UCLA	37.58	211.25	IQ composite (BSID-II, WPPSI, WISC,) Non-verbal IQ (MPSMT, LEITER) Receptive(RDLS, CELF-III) Expressive (RDLS, CELF-III) Adaptation-C (VABS) Adaptation-DLS(VABS) Adaptation-S (VABS) Adaptation composite (VABS)	3.63
Smith et al. (1997)	United States	Autism	90.48	36.95	27.57	Yes, R	11/10	UCLA	30.00	53.00	IQ composite (BSID-II)	2.90
Smith et al. (2000)	United States	Autism, PDD NOS	82.14	35.93	50.87	Yes, R	15/13	UCLA	24.52	250.67	IQ composite (BSID-II, SBIS) Non-verbal IQ (MPSMT) Receptive(RDLS) Expressive (RDLS) Language composite (RDLS) Adaptation-C (VABS) Adaptation-DLS(VABS) Adaptation-S (VABS) Adaptation composite (VABS)	2.90
Weiss (1999)	United States	Autism, PDD NOS	95.00	41.50	–	No	20	General	40.00	106.00	Adaptation composite (VABS)	2.02
Parent-managed intervention programs												
Anan et al. (2008) §	United States	Autism, PDD NOS	84.70	44.00	–	No	72	General	20.00	12.00	Receptive (MSEL) Expressive (MSEL) Adaptation-C (VABS) Adaptation-DLS (VABS) Adaptation-S (VABS) Adaptation-M (VABS) Adaptation composite (VABS)	3.55
Anderson et al. (1987)	United States	Autism, PDD NOS	76.92	43.00	57.83	No	13	General	20.00	53.00	IQ composite (BSID-II, SBIS) Language (PLS, SPT, PPVT, SICD) Adaptation composite (VABS)	2.53
Baker-Ericzen et al. (2007) §	United States	Autism, PDD NOS	83.00	49.36	–	No	158	Pivotal training	–	12.00	Adaptation-C (VABS) Adaptation-DLS (VABS) Adaptation-S (VABS) Adaptation composite (VABS)	3.88
Bibby et al. (2001)	United Kingdom	Autism	83.33	43.40	50.80	No	22	UCLA	5.85	31.60	IQ composite (WPPSI-R, WISC-III) Adaptation composite (VABS)	3.11
Reed et al. (2007) §	United Kingdom	Autism	100.00	41.89	53.40	No ‡	14	General	12.20	41.61	IQ composite (PEP-R) Adaptation composite (VABS)	2.81
Reed et al. (2007) §	United Kingdom	Autism	100.00	41.89	53.40	No ‡	13	General	27.00	41.61	IQ composite (PEP-R) Adaptation composite (VABS)	2.81
Sallows and Graupner (2005)	United States	Autism	80.00	34.20	52.10	No ‡	10	UCLA	31.28	198.85	IQ composite (BSID-II, WPPSI, WISC,) Non-verbal IQ (MPSMT, LEITER) Receptive (RDLS, CELF-III) Expressive (RDLS, CELF-III) Adaptation-C (VABS)	3.63

(continued on next page)

Table 1 (*continued*)

								ABA Intervention				
First author, year	Country	Diagnosis	Male (%)	Mean age (months)	Pre-IQ	Control group	Sample size*	Model	Intensity (h/week)	Duration (weeks)	Outcome (instrument)	Quality score†
Sheinkopf and Siegel (1998)	United States	Autism, PDD NOS	–	34.66	62.00	Yes, R	9/10	UCLA	19.45	68.90	Adaptation-DLS (VABS) Adaptation-S (VABS) Adaptation composite (VABS) Non-verbal IQ (MPSMT, CIIS, WPPSI, WISC)	2.54

*Total number of subjects for repeated-measures designs, or number of subjects in intervention/control groups for between-group studies.

†Quality score based on Downs and Black (1998) quality scale; rank: 0 (lowest quality) to 5 (highest quality).

‡Control and comparison groups in the studies by Sallows and Graupner (2005) and Reed et al., (2007) were analyzed separately as the control group received more than 10 weekly hours of ABA based intervention. IQ effect size for Birnbrauer and Leach (1993) was computed as a within-subject study as no post-test values are provided for the control group. Lovaas (1987) was analyzed as a within-subject study due to insufficient data reporting for the control group. Matos and Mustaca (2005) did not provide a standardized estimate of pre-intervention IQ.

§Studies not meeting the duration and intensity inclusion criteria but selected for meta-regression analyses.

Adaptation-C, Communication; Adaptation-DLS, Daily living skills; Adaptation-M, motor functioning; Adaptation-S, Socialization; BO, Systematic behavioral observation; BPLS-II, British Picture Language Scale (2nd Ed.); BSID-II, Bayley Scales of Infant Development (2nd Ed.); CELF-III, Clinical Evaluation of Language Fundamentals (3rd Ed.); CIIS, Cattell Infant Intelligence Scale; DP-II, Developmental profile II; LEITER, Leiter International Performance Scale; MPSMT, Merrill-Palmer Scales of Mental Tests; MSEL, Mullen Scales of Early Learning; PEP-R, Psycho-educational Profile (revised); PLS, Preschool Language Scale; PPVT, Peabody Picture Vocabulary Test; R, Randomized assignment; RDLS, Reynell Developmental Language Scales; REEL, Receptive-Expressive Emergence Language Scale; SBIS, Stanford-Binet Intelligence Scales; SBIS-IV, Stanford-Binet Intelligence Scales (4th Ed.); SICD, Sequenced Inventory of Communication Development; SPT, Symbolic Play Test; VABS, Vineland Adaptive Behavior Scales; WISC-III, Wechsler Intelligence Scale for Children (3rd Ed.); WPPSI-R, Wechsler Preschool and Primary Scale of Intelligence-Revised.

and end of the entire treatment period, even when the last follow-up outcome measure was reported in a separate paper.

For each outcome of interest, pooled estimates and 95% confidence intervals (CI) of effect sizes were calculated by using an inverse–variance weighted random-effects meta-analysis (Cottrell, Drew, Gibson, Holroyd, & O'Donnell, 2007). Between-study outcome variation (i.e., heterogeneity) was quantified with the I^2 statistic, which describes the percentage of variation across studies due to heterogeneity rather than chance regardless of treatment effect metric (Higgins & Thompson, 2002). Values around 25%, 50% and 75% refer to low, medium and high heterogeneity, respectively. Although I^2 was developed to be independent of the number of studies, it should be interpreted cautiously when few studies are meta-analyzed (Huedo-Medina, Sanchez-Meca, Marin-Martinez, & Botella, 2006).

When two or more studies were available, sensitivity analyses were performed by restricting the analysis to control group designs. In addition, separate meta-analyses were conducted by intervention model (UCLA, general ABA) and delivery format (clinic-based, parent-managed) to check consistency of treatment effects. At least two studies needed to be available for a sensitivity analysis to be conducted. For brevity, only effect size differences of 0.50 or above across ABA intervention models and intervention delivery format will be reported. Random-effects meta-regression (Thompson & Sharp, 1999) was used to separately evaluate whether results were different by population and intervention features, such as pre-intervention age, pre-intervention IQ, and treatment duration and intensity. For the purpose of analyzing the effects of intervention duration and intensity more thoroughly, studies were rank-ordered by total intervention hours (duration multiplied by intensity). A dose–response meta-analysis was conducted by studies' levels of total intervention hours. In order to strengthen the power of the analysis, studies excluded solely on the basis of limited treatment duration were included in the meta-regression and dose–response meta-analyses. Finally, publication and small-study effects biases were assessed using the extended Egger's test (Egger, Smith, Schneider, & Minder, 1997; Thompson & Sharp, 1999). Statistical analyses were carried out with Stata v. 8.1 (Stata Corporation, College Station, Texas).

3. Results

3.1. Study characteristics

Twenty-six studies met the pre-specified inclusion criteria (Fig. 1). Two studies were excluded because the relevant outcome was present in less than three papers (Boyd & Corley, 2001; Zachor, Ben Itzchak, Rabinovich, & Lahat, 2007). Two studies were excluded because of limited data reporting, including failure to provide pre-test measures and estimates of random variability (Luiselli, Cannon, Ellis, & Sisson, 2000; McEachin et al., 1993). The remaining 22 studies were included in the meta-analysis (Anderson, Avery, DiPietro, Edwards, & Christian, 1987; Ben, Lahat, Burgin, & Zachor, 2008; Ben-Itzchak & Zachor, 2007; Bibby, Eikeseth, Martin, Mudford, & Reeves, 2001; Birnbrauer & Leach, 1993; Cohen, Merine-Dickens, & Smith, 2006; Eikeseth, Smith, Jahr, & Eldevik, 2002; Eikeseth et al., 2007; Eldevik, Eikeseth, Jahr, & Smith, 2006; Harris & Handleman, 2000; Harris, Handleman, Gordon, Kristoff, & Fuentes, 1991; Howard, Sparkman, Cohen, Green, & Stanislaw, 2005; Lovaas, 1987; Magiati, Charman, & Howlin, 2007; Matos & Mustaca, 2005; Remington et al., 2007; Sallows & Graupner, 2005; Sheinkopf & Siegel, 1998; Smith, Eikeseth, Klevstrand, & Lovaas, 1997; Smith, Groen, & Wynn, 2000; Weiss, 1999). Three additional studies excluded solely on the basis of insufficient intervention duration were included in meta-regression and dose–response analyses (Anan, Warner, McGillivary, Chong, & Hines, 2008; Baker-Ericzen, Stahmer, & Burns, 2007; Reed,

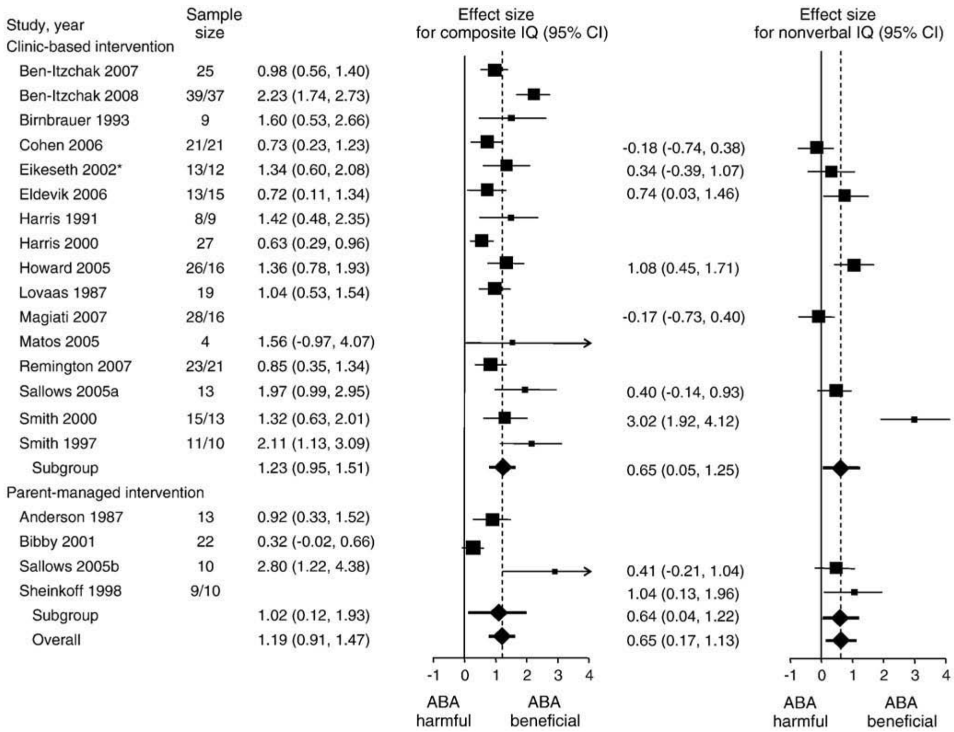

Fig. 2. Effect size for IQ and nonverbal IQ of applied behavior analysis intervention for participants with autism and pervasive developmental disabilities not otherwise specified. The area of each square is proportional to the study weight in the pooled analysis. Horizontal lines represent 95% confidence intervals (CI). Diamonds represent pooled estimates from inverse–variance weighted random-effects meta-analyses. Effect sizes and 95% CI are also presented numerically. Studies are classified by intervention delivery format (clinic-based, parent-managed). Sample sizes are total number of subjects for repeated-measures designs, or number of subjects in intervention/control groups for control group designs. Eikeseth et al., (2002) and Eikeseth et al., (2007) report data from the same cohort at different follow up periods; a single effect size was computed with the last follow up as post-intervention measure. Sallows and Graupner (2005) are reported as two independent repeated measures studies. Given that intervention and comparison groups at Sallows and Graupner (2005) received more than 10 weekly hours of intervention, each group was analyzed as an independent within subject study.

Table 2
Pooled effect sizes for IQ, language and adaptive behavior according to intervention features in studies of ABA intervention.[a]

Intervention feature	IQ composite			Language composite			Adaptive behavior composite		
	No. studies	Effect size (95% CI)	p value[b]	No. studies	Effect size (95% CI)	p value†	No. studies	Effect size (95% CI)	p value[b]
Pre-intervention age	20	−0.02 (−0.05 to 0.01)	0.157	5	−0.02 (−0.10 to 0.06)	0.621	18	0.01 (−0.04 to 0.06)	0.670
Pre-intervention IQ	19	0.00 (−0.03 to 0.03)	0.874	4	0.06 (−0.19 to 0.06)	0.317	14	0.03 (−0.02 to 0.07)	0.234
Duration, weeks	20	0.00 (−0.03 to 0.02)	0.500	5	0.01 (0.00 to 0.02)	0.001	18	0.00 (0.00 to 0.01)	0.346
Intensity, hours/week	19	0.01 (−0.01 to 0.04)	0.333	4	0.05 (−0.19 to 0.29)	0.705	17	0.05 (0.01 to 0.09)	0.015

CI, confidence interval.
[a] Pooled effect sizes were estimated from random-effects meta-regression models including indicator variables for each category of the intervention feature. Matos study was not included because only BSID-II raw scores were provided and in the absence of exact birth date, PDI scores under an IQ-equivalent scale could not be obtained.
[b] *p* value for heterogeneity of pooled effect sizes.

Osborne, & Corness, 2007). Two studies using control groups receiving more than 10 weekly hours of ABA intervention, were analyzed as independent repeated measures studies and will be referred to as separate studies (e.g., Reed et al., 2007; Sallows & Graupner, 2005). Two studies reporting data from the same cohort at different follow up periods (Eikeseth et al., 2002; 2007) were analyzed as a single study. In addition, Lovaas (1987) was analyzed as a within-subject study due to insufficient data reporting for the control group. The reader is referred to Table 1 for a systematic description of studies included in the meta-analysis. A summarization of study features is presented below.

The following outcomes were reported: full scale IQ (18 studies), nonverbal IQ (9 studies), receptive language (10 studies), expressive language (9 studies), language composite (5 studies), adaptive behavior–communication (10 studies), adaptive behavior–daily living skills (10 studies), adaptive behavior–socialization (10 studies), adaptive behavior–motor skills (3 studies), and overall composite adaptive behavior scores (14 studies). A complete listing of the instruments used to assess each of these outcomes is available in Table 1.

The mean quality score (of a possible maximum of 5) was 2.5 (range of 1.2 to 3.6). Studies tended to score higher in Reporting (0.8 out of 1.0) and Internal Validity-bias (0.7) as opposed to External Validity (0.4), Internal Validity-confounding (0.3) and Power (0.3). Quality scores by intervention model equaled 2.8 (range of 2.0 to 3.6) for UCLA model and 2.4 (range of 1.2 to 3.3) for general ABA intervention. Parent-managed programs obtained an average quality index of 3.0 (range of 2.5 to 3.6) while clinic-based programs scored 2.6 (range of 2.0 to 3.6). The reader is referred to Appendix A for the complete report of quality assessment.

A total of 323 subjects were included in intervention groups. The participants mean age ranged from 22.6 to 66.3 months. The percentage of male participants ranged from 55.6 to 97.7%. Fifteen studies reported results exclusively on children diagnosed with autism, while participants in 7 studies were both children diagnosed with autism and pervasive developmental disabilities not otherwise specified. With regard to intervention features, 13 studies followed the UCLA model, and 9 studies used the intervention model described as general ABA. Eighteen studies reported clinic- or school-based programs. Among them, two studies were delivered in the participants' home (Magiati et al., 2007; Weiss, 1999). Four trials reported data from parent-managed programs (Anderson et al., 1987; Bibby et al., 2001; Sallows & Graupner, 2005; Sheinkopf & Siegel, 1998). Intervention duration and intensity ranged from 48 to 407 weeks and from 12 to 45 weekly hours respectively.

There were 8 studies with within-subjects design included in the meta-analysis. Thirteen studies had control groups of which 6 used random or quasi-random assignment. Control groups of 9 studies comprised those having either an eclectic intervention or a combination of standard interventions including Treatment and Education of Autistic Children and related Communication Handicapped Children (TEACCH, see Piazza & Fadanni, 2002), special education classes and sensory integration therapy (Ben et al., 2008; Cohen et al., 2006; Eikeseth et al., 2002, 2007; Eldevik et al., 2006; Howard et al., 2005; Magiati et al., 2007; Sheinkopf & Siegel, 1998; Smith et al., 2000). One study used a public school special education group as control group (Remington et al., 2007) and in another study a group of typically developing children attending regular school participated as controls (Harris et al., 1991). The control group of the study by Smith et al. (1997) was comprised of children with autism receiving low intensity (i.e., <10 weekly hours) ABA intervention. Finally, Birnbrauer and Leach (1993) did not report any specific intervention in their control group.

3.2. Intelligence quotient

ABA intervention produced positive effects in all 18 studies reporting general IQ (Fig. 2). The pooled effect size across studies, covering a total of 311 participants, was 1.19 (95% CI 0.91 to 1.47, $p<0.001$). Effects tended to be stronger for clinic-based programs compared to parent-managed interventions with effect sizes of 1.23

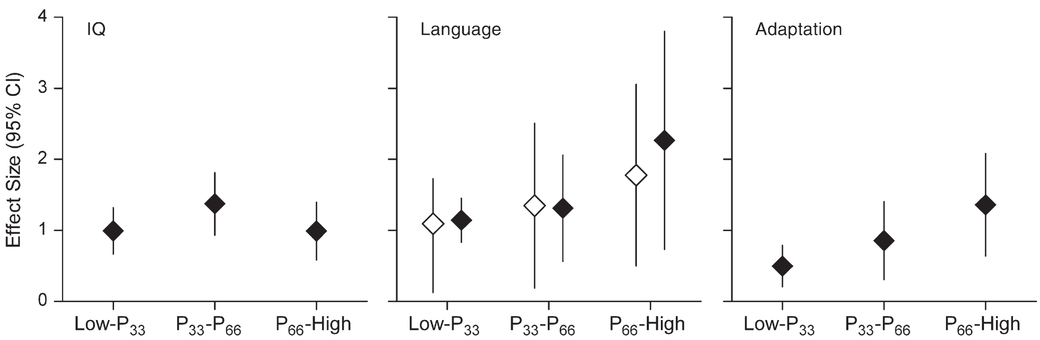

Fig. 3. Dose–response meta-analysis by levels of applied behavior analysis total intervention hours for IQ, language (receptive, expressive) and adaptive behavior (composite scores). Total intervention hours levels were established by percentile 33 (P_{33} = 1833.8) and 66 (P_{66} = 4129.3) of total intervention hours of all meta-analyzed studies (treatment intensity multiply by duration). Open diamonds in the central graph show expressive language and solid diamonds show receptive language. Horizontal lines represent 95% confidence intervals (CI). Diamonds represent pooled estimates from inverse–variance weighted random-effects meta-analyses by total intervention hour levels.

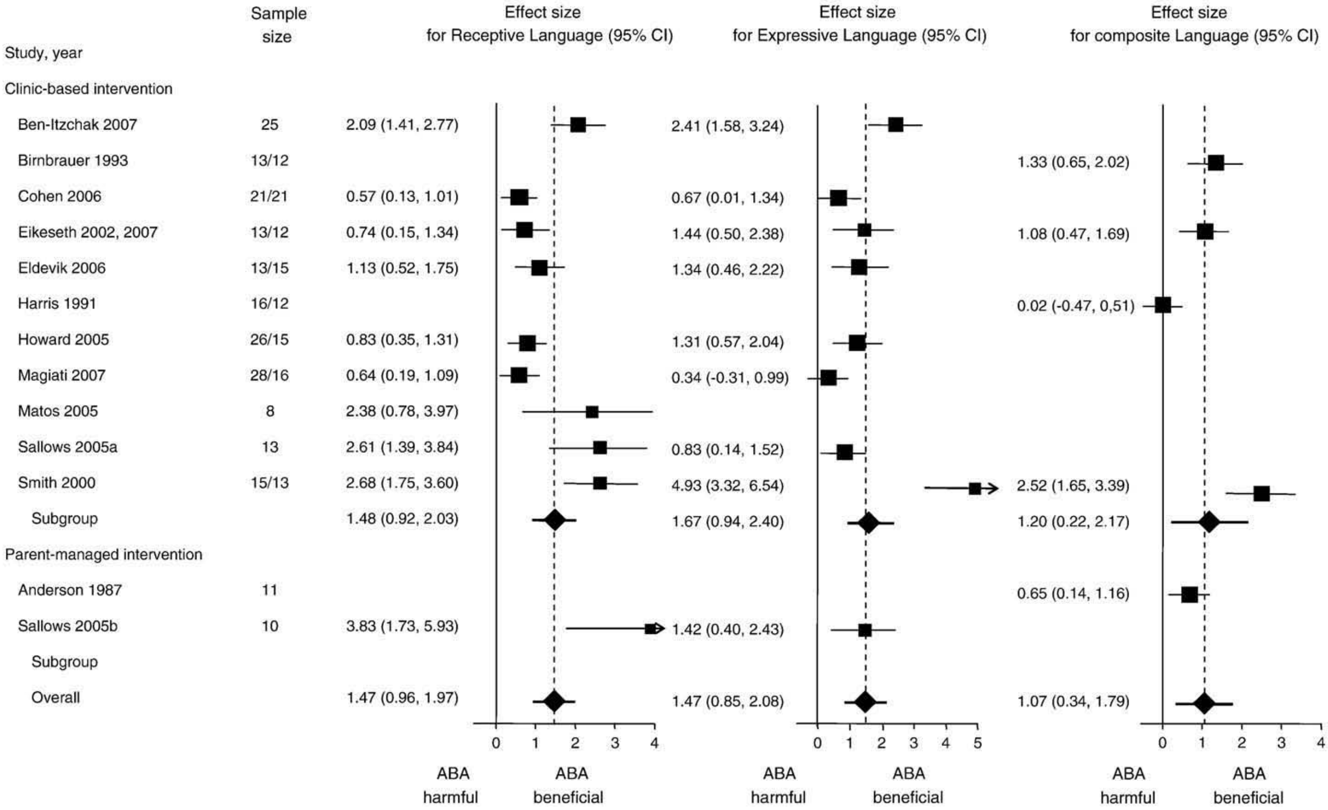

Fig. 4. Effect size for receptive language, expressive language and composite language of applied behavior analysis intervention for participants with autism and pervasive developmental disabilities not otherwise specified. See Fig. 2 notes.

(95% CI 0.95 to 1.51, $p<0.001$) versus 1.02 (95% CI 0.12 to 1.93, $p<0.027$) (see Fig. 2). Similar results were obtained when analysis was restricted to the 10 studies ($n=169$) that included a control group with an effect size of 1.31 (95% CI 0.92 to 1.70, $p<0.001$). Meta-regression did not show clear effects of intensity or duration (Table 2). Dose–response meta-analysis of studies' total treatment duration suggested that high total treatment duration did not improve treatment gains above average levels (Fig. 3). There was evidence of heterogeneity ($I^2=75\%$, 95% CI 60 to 84%) and publication bias ($p=0.012$).

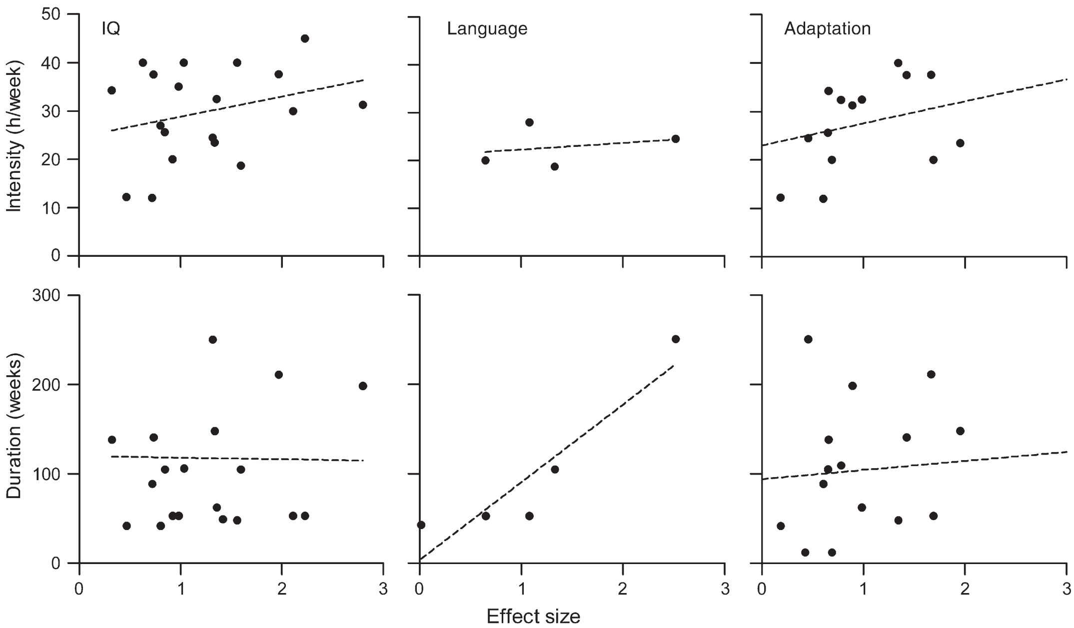

Fig. 5. Scatter plot and linear regression line of intervention intensity and duration by intervention effect size for IQ, language (composite score) and adaptive behavior (composite score) outcomes.

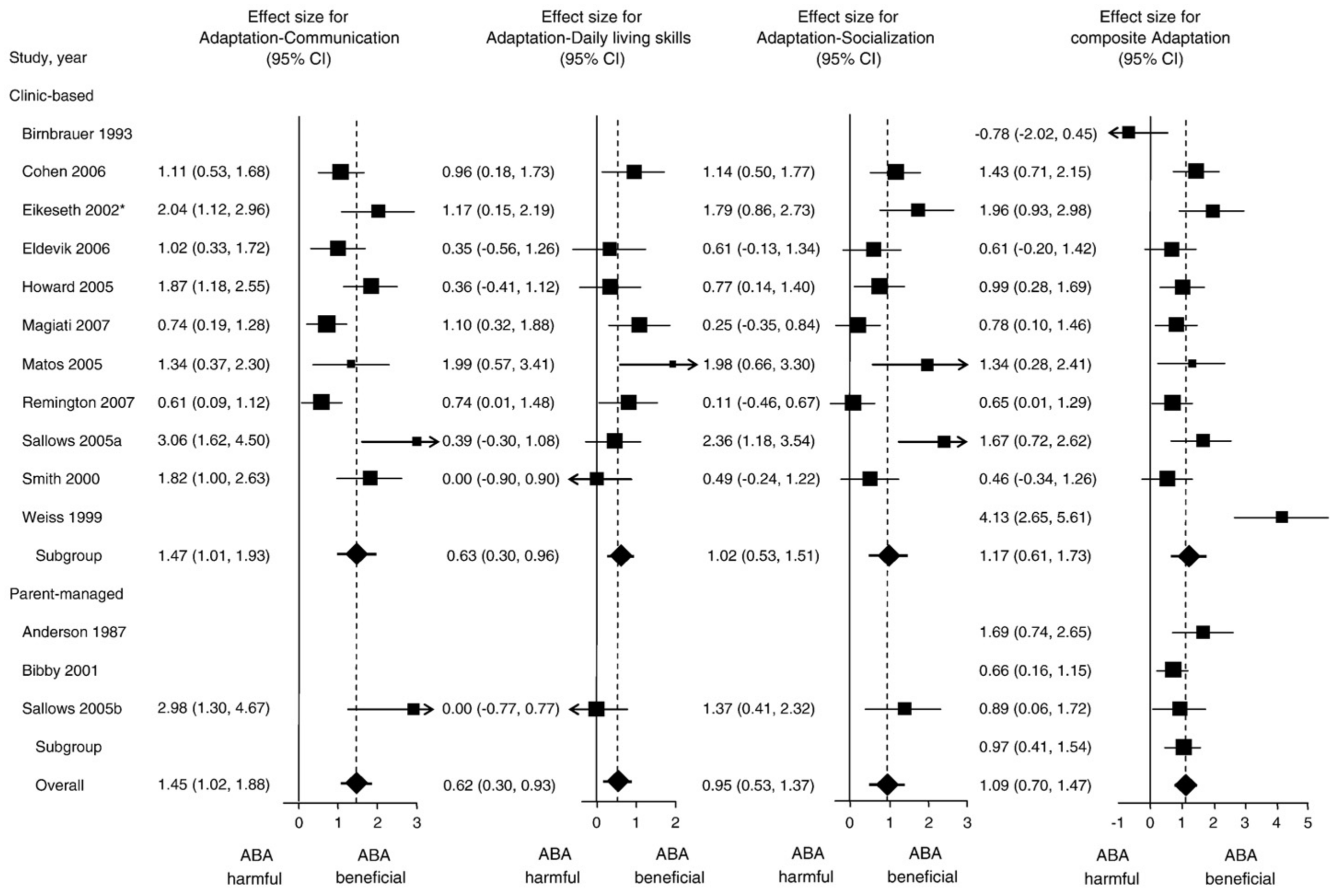

Fig. 6. Effect size for adaptive behavior domains including communication, daily living skills, socialization and adaptive behavior composite scores of applied behavior analysis intervention for participants with autism and pervasive developmental disabilities not otherwise specified. See Fig. 2 notes.

Non-verbal IQ was reported by 10 studies for a total of 146 participants. All but one reported positive effects of ABA intervention. The pooled effect size for this outcome was 0.65 (95% CI 0.17 to 1.13, $p=0.008$). Effects were similar across intervention models: clinic-based programs showed an effect size of 0.65 (95% CI 0.05 to 1.25, $p=0.033$) and the parent managed programs effect size was 0.65 (95% CI 0.05 to 1.25, $p=0.034$). Eight of these studies included a control group covering 123 participants. When meta-analysis was restricted to controlled studies the effect size was 0.76 (95% CI 0.10 to 1.42, $p=0.024$). There was evidence of heterogeneity ($I^2=78\%$, 95% CI 59 to 88%) and publication bias ($p=0.013$).

3.3. Language skills

Receptive language was assessed in 11 different studies that provided ABA intervention to 172 participants. All of these studies reported favorable effects of the intervention for receptive language performance with a pooled effect size of 1.48 (95% CI 0.96 to 1.97, $p<0.001$) (Fig. 4). When meta-analysis was restricted to the 7 studies that included a control group ($n=116$) then an effect size of 0.99 (95% CI 0.56 to 1.42, $p<0.001$) was obtained. There was evidence of heterogeneity ($I^2=81\%$ CI 65 to 89%). Egger's test suggested publication bias ($p=0.048$).

Expressive language was reported in 10 studies covering 164 participants. All studies reported favorable effects of ABA intervention for expressive language skills as measures by standardized assessments. Pooled random meta-analysis effect size was 1.47 (95% CI 0.85 to 2.08, $p<0.001$). There was evidence of heterogeneity (I^2 = 80%, CI 62 to 89%) and publication bias was likely ($p=0.003$). Both results for receptive and expressive language demonstrated clear dose–response trends for intervention total duration (Fig. 3).

General language skills were reported in 5 studies that provided ABA intervention to 64 participants with a pooled effect size of 1.07 (95% CI 0.34 to 1.79, $p=0.004$). All but one of these studies showed distinctively favorable effects of ABA intervention (Fig. 4). When meta-analysis was restricted to the 4 studies that included a control group, effect size equaled 1.20 (95% CI 0.22 to 2.17, $p=0.017$). Meta-regression indicated that effect size increased directly with intervention duration (Fig. 5, Table 2). There was evidence of heterogeneity ($I^2=86\%$ 95% CI 69 to 94%) and publication bias ($p=0.009$).

3.4. Adaptive behavior domains

Adaptive behavior was assessed by means of standardized assessments of competence in the domains of communication, daily living skills, motor skills, socialization and an adaptive behavior composite measure. Communication, daily living skills and socialization were assessed in 11 studies ($n=170$). All studies reported favorable effects of ABA intervention for these three domains (Fig. 6). Communication resulted in a pooled effect size of 1.45 (95% CI 1.02 to 1.88, $p<0.001$) with all studies reporting favorable effect sizes. Sensitivity analysis restricted to the 8 controlled studies ($n=138$) provided an effect size of 1.25 (95% CI 0.83 to 1.67, $p<0.001$). The effect size for communication tended to be higher for the 6 studies implementing the UCLA model intervention (ES = 1.73, 95% CI 1.06 to 2.39, $p<0.001$) as opposed to the 4 studies implementing general ABA intervention (ES = 1.17, 95%CI 0.59 to 1.76, $p<0.001$). There was evidence of heterogeneity ($I^2=68\%$ CI 38 to 83%) and publication bias ($p=0.002$). Effect size for daily living skills reached 0.62 (95% CI 0.30 to 0.93, $p<0.001$), whereas meta-analysis restricted to the 8 studies including a control group was 0.68 (95% CI 0.36 to 0.99, $p<0.001$). There was no evidence of heterogeneity ($I^2=27\%$ 95% CI 0 to 65%) or publication bias ($p=0.191$). Socialization produced a pooled effect size of 0.95 (95% CI 0.53 to 1.37, $p<0.001$). Sensitivity analysis restricted to the 8 controlled studies ($n=138$) resulted in an effect size of 0.68 (95% CI 0.29 to 1.06, $p=0.001$). There was strong evidence of heterogeneity ($I^2=66\%$ 95% CI 34 to 83%) and publication bias ($p=0.002$). Motor skills data were reported in three controlled studies for 51 participants. All three studies reported positive effects of ABA intervention when examined individually. The pooled effect size was 0.71 (95% CI 0.19 to 1.22, $p=0.008$). There was no evidence of heterogeneity ($I^2=0\%$ CI 0 to 90%) or publication bias ($p=0.109$). An adaptive behavior composite measure (combining all four domains described above) was reported in 15 distinct papers ($n=232$). Thirteen out of these 15 studies showed positive effects of ABA intervention (Fig. 6). The pooled effect size was 1.09 (95% CI 0.70 to 1.47, $p<0.001$). Results were consistent across program delivery formats; clinic-based programs had an effect size of 1.17 (95% CI 0.70 to 1.47, $p<0.001$) and parent-managed programs had a pooled effect size of 0.97 (95% CI 0.61 to 1.739, $p=0.001$). Meta-analysis limited to the 10 studies that included a control group ($n=165$) produced a somewhat smaller effect size (0.81; 95% CI 0.39 to 1.23, $p<0.001$). Effects increased with intervention intensity while duration did not affect effect size (Table 2). Dose–response meta-analyses demonstrated a clear increase in effect sizes by treatment total duration (Fig. 3). There was a strong heterogeneity effect ($I^2=68\%$ 95% CI 44 to 82%), while publication bias was not apparent ($p=0.091$).

4. Discussion

The overall appreciation of long-term, comprehensive ABA intervention effects for autism through standardized molar skills assessments has been hampered by the varying methods, designs and treatment features of published studies. In an attempt to fill this gap in the literature, state-of-the-art meta-analytical methods were implemented, including quality assessment, sensitivity analyses, meta-regression, dose–response meta-analysis and meta-analysis of studies of different metrics. Results suggest that long-term, comprehensive ABA intervention leads to (positive) medium to large effects in terms of intellectual functioning, language development, and adaptive behavior of individuals with autism. Although favorable effects were apparent across all outcomes, language-related outcomes (IQ, receptive and expressive language, communication) were distinctively superior to non-verbal IQ, social functioning and daily living skills, with effect sizes approaching 1.5 for receptive and expressive language and communication skills. This is particularly noteworthy as qualitative impairments in communication are one of the core features of autism. This finding is also consistent with the amount of time devoted by most ABA curricula to language and communication skills (Maurice et al., 1996). A trend strengthened in recent years through the development of novel ABA intervention procedures for language (Greer & Ross, 2007; Sundberg, 2007).

Pooled effects were consistent when analysis was restricted to controlled studies. In addition, potential confounding sources as pre-intervention age and pre-intervention IQ did not make a difference to treatment effectiveness (Table 2). Effects were also consistent for both clinic-based and parent-managed programs with slightly superior effect sizes found for clinic-based programs in terms of IQ, nonverbal IQ and adaptive behavior composite measures. Meta-analysis of all outcomes other than composite language skills and motor skills showed similar effects of both ABA intervention models (i.e., UCLA, general ABA) for all outcomes with the exception of communication, which showed stronger positive effects for the UCLA based programs. However, as these differential effects were not substantiated by expressive and receptive language outcomes, the significance of this finding remains unclear.

Although the comparison between ABA intervention models may be highly informative, more sophisticated distinctions including operational definitions of each intervention model and the addition of intervention fidelity measures should be employed to help discriminate models. Fidelity measures and standards cannot currently be assumed for studies in this field (McIntyre, Gresham, DiGennaro, & Reed, 2007; Wheeler, Baggett, Fox, & Blevins, 2006). The adoption of these quality standards would help to interpret any model-specific effects to be found in the future in terms of the curricula, programs and behavioral mechanisms that may be distinctive of any particular approach. On the other hand, this will also help professionals and clients to determine what these approaches may have in common.

Meta-regression analysis provided a clear account of the impact of intervention intensity and duration that is not obvious from the simple examination of individual studies. Overall language skills tended to benefit more from intervention duration while functional and psychosocial adaptive behaviors benefited more from intervention intensity. When meta-analysis was replicated for levels of total intervention duration, dose–response effects were evident for language performance and functional and psychosocial adaptive behaviors, while dose–response analysis for intellectual functioning showed, to some extent, an exhaustion of intervention effects. The highest magnitude of dose–response effects were demonstrated for receptive and expressive language. This finding, combined with the strong effects reported for language-related outcomes, suggest that verbal repertoires have a great potential for continuous treatment gains as opposed to other repertoires that may follow an asymptotic profile. Exhaustion of treatment effects by increasing levels of treatment intensity have been suggested before (Reed et al., 2007), however, our results indicate that this pattern may be different for intellectual functioning, verbal skills and functional and psychosocial repertoires.

Inclusion of repeated measures studies provides a preliminary external verification of controlled studies effects, particularly for those studies that did not have no-treatment controls, did not used randomization and reported small sample sizes. Namely, consistency of treatment effects across within-subjects and controlled studies strengthen the plausibility that control groups' composition was not severely biased and did not affect treatment outcomes to a high extent. Similarly, consistency of treatment effects across within-subjects and controlled studies provides also and external indication of within-subjects studies internal validity. Namely, consistency of effects across study designs suggests that within-subjects studies were not severely affected by design-specific threats including trend in baseline and over-reported effect sizes due to smaller variability.

As control groups were generally those receiving eclectic interventions (e.g., special education, sensory integration, TEACCH and others), meta-analysis also provides a preliminary comparison between ABA intervention and other forms of treatment for autism. This is an interesting extension of this study as there are few formal comparisons of ABA intervention effects to other treatment paradigms (Delprato, 2001; Reed et al., 2007). However, this comparison is only tentative; a formal comparison of intervention paradigms would require that two or more intervention groups have comparable treatment intensity and equal treatment fidelity requirements, which was not the case for the meta-analyzed studies herein reported. Nonetheless, the results of this meta-analysis are straightforward in their current form. Therefore, the superiority of ABA intervention suggested by these data shall not be discounted.

Randomization to group assignment was seldom implemented in the studies found, and the use of quasi-random assignment strategies (e.g., assignment to control or experimental group depending upon therapists' availability) raises various ethical and internal validity concerns. General quality standards of clinical studies including randomization, blindness, intention to treat analysis, and the use of prospective (as opposed to retrospective) designs, were inconsistently observed. In addition, quality standards specific to this field, e.g. comparable pre-intervention IQ across groups and treatment fidelity standards, were generally not followed (McIntyre et al., 2007; Wheeler et al., 2006). Although random effects meta-analysis and sensitivity analysis may partially compensate for this deficit, somewhat different results might well be found if studies employ such higher methodological standards. Moreover, publication bias was evident in all outcomes but daily living skills, motor functioning and composite adaptation. However, the limited sample size of most studies suggests that evidence of publication bias may simply be the byproduct of small sample size studies rather than genuine publication bias (Whitehead, 2002).

Recommendations for clinician and researchers planning to do controlled studies in this area include: (1) the observation of clinical trials quality standards including intention to treat analysis and randomization (see CONSORT guidelines for a complete listing of quality standards; Moher, Schulz, Altman, & the CONSORT Group, 2001), (2) use no-treatment controls or match treatment intensity and duration across groups, (3) monitor the degree to which therapist adhere to treatment protocols in the intervention group and also in the comparison group whenever controls follow an alternative treatment, (4) implement specific approaches to treatment in order to provide direct comparisons of different intervention paradigms both within ABA intervention (e.g., UCLA, CABAS, verbal behavior) and between ABA intervention and other forms of treatment (e.g., pharmacological, TEACCH).

A wide adoption of these standards may establish a clearer picture of the highly promising effects of ABA intervention and may constitute the basis for decision-making in public health and social policies relating to autism and developmental disabilities.

Acknowledgements

The author is in debt with Dr. Neil Martin and Dr. Stephen N. Haynes for their useful comments to earlier versions of this manuscript. Mónica Rodríguez-Mori replicated the quality assessment. The author is also grateful to Ming Tang for providing Chinese to English translations during the screening of references.

This study has not been funded nor subject to pre-submission approval by any institution including the ones to which the author is affiliated. The contents expressed in this article do not necessarily represent the position of the institutions to which the author is affiliated.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.pscychresns.2010.02.004.

References[1]

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders 4th Edition, Text Revised.* Arlington VA: American Psychiatric Publishing, Inc.

*Anan, R. M., Warner, L. J., McGillivary, J. E., Chong, I. M., & Hines, S. J. (2008). Group Intensive Family Training (GIFT) for preschoolers with autism spectrum disorders. *Behavioral Interventions*, *23*, 165−180.

*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.

*Baker-Ericzen, M. J., Stahmer, A. C., & Burns, A. (2007). Child demographics associated with outcomes in a community-based pivotal response training program. *Journal of Positive Behavior Interventions*, *9*, 52−60.

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, *41*, 257−278.

Ben-Itzchak, I. E., Lahat, E., Burgin, R., & Zachor, A. D. (2008). Cognitive, behavior and intervention outcome in young children with autism. *Research in Developmental Disabilities*, *29*, 447−458.

*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. (2001). Progress and outcomes for children with autism receiving parent-managed intensive interventions. *Research in Developmental Disabilities*, *22*, 425−447.

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

Botella, J., & Gambara, H. (2006). Doing and reporting meta-analysis. *International Journal of Clinical and Health Psychology*, *6*, 425−440.

*Boyd, R. D., & Corley, M. J. (2001). Outcome survey of early intensive behavioral intervention for young children with autism in a community setting. *Autism*, *5*, 430−441.

British Columbia Office of Health Technology Assessment. (2001). Autism and Lovaas treatment: A Systematic review of effectiveness evidence. *International Journal of Technology Assessment in Health Care, 17*, 252. Complete reference list retrieved May 13, 2009, from http://www.chspr.ubc.ca/files/publications/2000/bco00-01T_autism3.pdf

*Cohen, H., Merine-Dickens, M., & Smith, T. (2006). Early intensive behavioral treatment: replication of the UCLA model in a community setting. *Development and Behavioral Pediatrics*, *27*, S145−S155.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis*, 2nd ed. Upper Saddle River, NJ: Pearson.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). Definition and characteristics of applied behavior analysis. *Applied Behavior Analysis* (pp. 2−23)., 2nd ed. Upper Saddle River, NJ: Pearson.

Cottrell, C., Drew, J., Gibson, J., Holroyd, K., & O'Donnell, F. (2007). Feasibility assessment of telephone-administered behavioral treatment for adolescent migraine. *Headache: The Journal of Head and Face Pain*, *47*, 1293−1302.

Delprato, D. J. (2001). Comparisons of discrete-trial and normalized behavioral intervention for young children with autism. *Journal of Autism and Developmental Disorders*, *31*, 315−325.

Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality of both randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health*, *53*, 377−384.

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 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 and Developmental Disorders*, *36*, 211−224.

Eldevik, S., Hastings, R. P., Hughes, J. C., Jahr, E., & Eikeseth, S. (2009). Meta-analysis of early intensive behavioral intervention for children with autism. *Journal of Child and Adolescent Psychology*, *38*, 439−450.

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 and Intervention in Developmental Disabilities*, *5*, 49−58.

Foxx, R. M. (2008). Applied behavior analysis treatment of autism: The state of the art. *Child and adolescent psychiatric clinics of North America*, *17*, 821−834.

Greer, R. D., & Ross, D. E. (2007). *Verbal behavior analysis: Inducing and expanding new verbal capabilities in children with language delays.* Upper Saddle River, NJ: Allyn & Bacon.

*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.

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 behavior analytic and eclectic treatments for young children with autism. *Research in Developmental Disabilities*, *26*, 359−383.

Huedo-Medina, T. B., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in Meta-Analysis: Q test or I^2 Index? *Psychological Methods*, *11*, 193−206.

Hughes, J. R. (2008). A review of recent reports on autism: 1000 studies published in 2007. *Epilepsy & Behavior*, *13*, 425−437.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24*, 659−685.

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (1999). *Autism Diagnostic Observation Schedule.* Los Angeles: Western Psychological Services.

Lovaas, O. I. (1981). *Teaching developmentally disabled children: The me book.* Baltimore: University Park Press.

*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., Ellis, J. T., & Sisson, R. W. (2000). Home-based behavioral interventions for young children with autism/pervasive developmental disorder: A

[1] References marked with an asterisk indicate studies included in the meta-analysis.

preliminary evaluation of outcome in relation to child age and intensity of service delivery. *Autism*, *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.
*Matos, M. A., & Mustaca, A. E. (2005). Análisis Comportamental Aplicado (ACA) y Trastornos Generalizados del Desarrollo (TGD): Su evaluación en Argentina [Applied behavior analysis and pervasive developmental disabilities: Assessment in Argentina]. *Interdisciplinaria*, *22*, 59−76.
Maurice, C., Green, G., & Luce, S. C. (Eds.). (1996). *Behavioral intervention for young children with autism: A manual for parents and professionals*. Austin, TX: Pro-Ed, Inc.
Maurice, C., Green, G., & Foxx, R. M. (2001). *Making a difference: Behavioral intervention for autism.* Austin, TX: Pro-Ed, Inc.
*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.
McIntyre, L. L., Gresham, F. M., DiGennaro, F. D., & Reed, D. D. (2007). Treatment integrity of school-based interventions with children in the Journal of Applied Behavior Analysis 1991–2005. *Journal of Applied Behavior Analysis*, *40*, 659−672.
Moher, D., Schulz, K. F., & Altman, D.the CONSORT Group. (2001). The CONSORT Statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *Journal of the American Medical Association*, *285*, 1987−1991.
Morris, S. B. (2008). Estimating effect sizes from pretest–posttest-control group designs. *Organizational Research Methods*, *11*, 364−386.
Morris, S., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*, 105−125.
New York State Department of Health. (1999). *Clinical practice guideline: Report of the recommendations. Autism/pervasive developmental disorders. Assessment and intervention for young children (age 0-3 years)* (Publication No. 4215). Albany, NY: Author. Retrieved January 1, 2009, from http://www.health.state.ny.us/community/infants_children/early_intervention/autism/9
Ontario Ministry of Education. (2007). *Questions and Answers: Students with autism spectrum disorders.* Toronto, Ontario: Author. Retrieved May 13, 2009, from http://www.edu.gov.on.ca/eng/general/elemsec/speced/autism.html
Piazza, S. D., & Fadanni, D. (2002). Autism, TEACCH and Lovaas methods: Towards a synthesis. *Revue Medicale de Liege*, *57*, 463−466.
*Reed, P., Osborne, L. A., & Corness, M. (2007). Brief report: Relative effectiveness of different home-based behavioral approaches to early teaching intervention. *Journal of Autism and Developmental Disorders*, *37*, 1815−1821.
*Remington, B., Hastings, R. P., Kovshoff, H., degli 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.
Robinson, K. A., & Dickersin, K. (2002). Development of a highly sensitive search strategy for the retrieval of reports of controlled trials using PubMed. *International Journal of Epidemiology*, *31*, 150−153.
*Sallows, G. O., & Graupner, T. D. (2005). Intensive behavioral treatment for children with autism: Four-year outcome and predictors. *American Journal of Mental Retardation*, *110*, 417−438.
Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing single-subject research. *Behavior Modification*, *22*, 221−242.
Severtson, J. M., Carr, J. E., & Lepper, T. L. (2009). Noncontingent reinforcement is an empirically supported treatment for problem behavior exhibited by individuals with developmental disabilities. *Research in Developmental Disabilities*, *30*, 44−57.
*Sheinkopf, S. J., & Siegel, B. (1998). Home based behavioral treatment of young children with autism. *Journal of Autism and Developmental Disorders*, *28*, 15−23.
Shook, G. L. (2005). An examination of the integrity and future of the Behavior Analyst Certification Board credentials. *Behavior Modification*, *29*, 562−574.
*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 of Mental Retardation*, *105*, 269−285.
Sundberg, M. (2007). Verbal behavior. In J. O. Cooper, T. E. Heron, & W. Heward (Eds.), *Applied behavior analysis* (pp. 525−547)., 2nd ed. Upper Saddle River, NJ: Pearson.
Thompson, S. G., & Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: A comparison of methods. *Statistics in Medicine*, *18*, 2693−2708.
U.S. Public Health Service. (1999). Mental health: A report of the surgeon general (chap. 3, section on autism). Retrieved May 13, 2009, from http://www.surgeongeneral.gov/library/mentalhealth/chapter3/sec6.html#autism
Volkmar, F. R., & Davies, M. (2003). Autism. *Lancet*, *362*, 133−141.
*Weiss, M. J. (1999). Differential rates of skill acquisition and outcomes of early intensive behavioral intervention for autism. *Behavioral Interventions*, *14*, 3−22.
Wheeler, J. J., Baggett, B. A., Fox, J., & Blevins, L. (2006). Treatment integrity: A review of intervention studies conducted with children with autism. *Focus on Autism and Other Developmental Disabilities*, *21*, 45−54.
Whitehead, A. (2002). *Meta-analysis of controlled clinical trials.* Hoboken, NJ: John Wiley and Sons.
*Zachor, D. A., Ben Itzchak, E., Rabinovich, A., & Lahat, E. (2007). Change in autism core symptoms with intervention. *Research in Autism Spectrum Disorders*, *1*, 304−317.