This is a post print version of the manuscript. Please refer to the published article for the final version.

Assessing Spoken Language Outcomes in Children with ASD: A Systematic Review

David Trembath

Menzies Health Institute Queensland, Griffith University, Australia

Marleen Westerveld

Griffith Institute of Education Research, Griffith University, Australia

Leanne Shellshear

School of Allied Health Sciences, Griffith University, Australia

Key words: Comprehensive intervention; early intensive behavioural intervention (EIBI); language; Autism Spectrum Disorder

Correspondence:

David Trembath

Menzies Health Institute Queensland

Griffith University

Ph: +61 7 5678 0103

E: D.Trembath@Griffith.edu.au

Abstract

Comprehensive early intervention programs based on the principles of applied behavior analysis can promote social-communication development in children with Autism Spectrum Disorder (ASD). However, it is not clear to what extent these positive results represent progress towards helping children achieve the ultimate goal of spontaneous, flexible, directed, and effective spoken language for a range of purposes across all life contexts. The aim of this systematic review was to document the intervention targets and assessment tools used to measure spoken language outcomes for children with ASD receiving comprehensive intervention programs. A total of 23 studies met the inclusion criteria. Although social-communication functioning was the most frequently targeted aspect of development, assessment of these skills relied predominantly on the use of standardised assessments. No studies included sampling and analysis of the children's spoken language in daily activities. The lack of comprehensive measures with high ecological validity limits our ability to draw conclusions regarding spoken language outcomes from the studies to date.

Introduction

The ultimate communication goal for all children in the preschool years is the development and use of spontaneous, flexible, directed, and effective spoken language (or multimodal communication for some children with physical disability) for a range of purposes across all life contexts. This requires the coordinated acquisition of a sophisticated range of receptive and expressive language skills across the domains of morphology (word structure), syntax (sentence structure), semantics (meaning), phonology (organisation of sounds), and pragmatics (interaction) [1]. For children with Autism Spectrum Disorder (ASD) who, by definition, have pervasive impairments in social-communication skills and behaviour, the path to spoken language development is compromised [2]. Accordingly, a range of evidence-based

interventions have been developed with some promising outcomes for communication reported [3]. In this article, we review the skills targeted and assessment tools used to measure intervention outcomes for children with ASD, and consider the suitability of these tools for measuring changes in spoken language.

Comprehensive versus Focused Interventions

Odom, Collet-Klingenberg [4] used the terms ‘comprehensive treatment models’ (CTMs) and ‘focused intervention practices’ (FIPs) to categorise the range of interventions developed for children with ASD. Comprehensive treatment models comprise a set of practices designed to address the core deficits of ASD across multiple developmental domains, often delivered through a manualised program [e.g., Early Start Denver Model; 5]. Focused intervention practices include a set of targeted strategies designed to change a specific behaviour or to achieve a specific developmental outcome [e.g., The Picture Exchange Communication System (PECS) targeting requesting, commenting, and response to questions; 6]. Based on a systematic review of research evidence, Maglione, Gans, Das et al., [7] concluded that children with ASD should have access to intensive comprehensive interventions, which may also be supplemented with focused interventions to meet their individual needs. Here, we focus on comprehensive early intensive interventions based on the principles of applied behavior analysis.

Evidence for Comprehensive Interventions

There is growing evidence that comprehensive interventions can facilitate receptive and expressive language development in children with ASD. Reichow, Barton, Boyd, and Hume [8], for example, completed a Cochrane review of five studies in which Early Intensive Behavioral Interventions (EIBIs) were compared to treatment as usual in the community for 203 children with ASD under 6 years of age. Based on the principles of applied behavior analysis, EIBIs target multiple developmental domains and are delivered to young children at an intensity of between

20-40 hours per week. Compared to 'treatment as usual,' Reichow et al., reported that EIBIs resulted in superior gains in adaptive behavior, social skills, intelligence, communication and language, ASD symptoms, and quality of life compared to treatment as usual in the community. However, they noted that only a small number of studies were of sufficient scientific quality to warrant inclusion, and called for further research. Other researchers [e.g., 9] have observed that even within well-controlled studies yielding positive group results, considerable individual variability in outcomes is common. This lack of high quality research studies and variability in child outcomes presents a major challenges for parents, educators, and researchers when attempting to infer the likely benefits of an intervention for a particular child, including the possible impact on his or her receptive and expressive language development [10].

The challenge of assessing outcomes

The challenge of interpreting findings is compounded by the selection of skills to be targeted (i.e., dependent variables) and the manner in which outcomes are measured and reported in intervention studies. Parents, educators, researchers, and policy makers require data collected from outcome measures that are sensitive, objective, and ecologically and socially valid in order to judge the effectiveness of a particular intervention [11]. However, there is evidence to suggest that such measures are not always used. Matson and Rieske [12], for example, documented outcome measures used in EIBI research published from 1987 to 2013. Of the 25 studies presented, 22 utilised standardised assessments to evaluate changes in intellectual functioning and adaptive behavior. Five studies included standardised assessment of speech and language development. Although standardised assessments are an essential tool for evaluating outcomes [13], there are several challenges associated with their use with children with ASD. Matson [14] noted that the commonly used Vineland Adaptive Behavior Scales [15] was not designed for use with very young children with ASD presenting with complex learning and language difficulties, resulting in an artificial ceiling effect leading to low capacity to discriminate differences in skill

level. Furthermore, it is likely that when assessing young children with ASD, changes in IQ or adaptive behavior over time may often reflect changes in compliance and test-taking ability [14].

There is also concern that many commonly used standardised assessment tools may not adequately capture children's level of functioning in everyday activities. Neisworth and Bagnato [16], for example, argued that standardised assessments conducted in highly controlled clinical settings with strict administration guidelines are unlikely to provide an accurate representation of each child's strengths and difficulties across a range of settings. McConachie et al., [11], following their systematic review of assessment tools used in ASD research, concluded that (a) only a small number of tools had been shown to be valid measures for children with ASD and (b) that the information provided by commonly used assessment tools aligned poorly with the information parents reported to be valuable in judging outcomes. With regards to spoken language, the parent-report McArthur Bates Communicative Development Inventories (MB-CDI) [17] was the only assessment tool that featured in the list of 12 tools presented in the 'summary of quality tools' indicating that it met criteria for three or more quality indicators used in the review (i.e., used in high quality study, blinding of assessors, can be used for children in the 0-6 years range, and had been used by more than one research group). While the MB-CDI has excellent face validity, provides a detailed measure of vocabulary, is relatively quick to administer, and has been shown to be a sensitive measure of changes in child language [e.g., 18, 19], it does not provide an objective and comprehensive measure of children's spoken language outcomes. McConachie et al. [11] also noted that although parent-report measures have excellent external validity, they are at greater risk of bias than clinician-delivered direct assessments. These findings illustrate the challenge of selecting appropriate assessment tools in ASD research, and the need for parents, educators, and researchers to not only critically appraise the findings of research, but also the assessment tools from which the outcome data have been derived.

Measuring Spoken Language Outcomes

Given that supporting children to develop spoken language is a key goal of intervention, and considering the concerns regarding commonly used assessment tools, it seems prudent to examine the assessment tools commonly used in interventions studies. But what might be a valid measure of spoken language? Key practice guidelines [e.g., 20, 21, 22] for the assessment of communication skills in individuals at risk for, or diagnosed with ASD, include a thorough examination of (a) communicative forms (e.g., gestures, vocalisations, eye gaze, physical actions, use of augmentative communication, idiosyncratic behaviours including behaviours of concern); (b) communicative functions (e.g., sharing information, requesting, protesting); (c) frequency, effectiveness, and social quality of verbal and non-verbal communication strategies; (d) the extent to which communication strategies are coordinated; (e) and the use of atypical communication strategies. While standardised assessments form an essential part of the recommended battery for assessing this broad range of skills, they do not necessarily provide a comprehensive picture of each child's skills and needs in everyday settings.

Kasari et al [23], following a review of assessment tools for minimally verbal children with ASD, noted that "...most of the measures have serious limitations for use with minimally verbal children, which have severely impeded progress in both research and clinical practice" (p. 12). They recommended a comprehensive approach to measuring children's communication skills and outcomes, including the collection and analysis of language samples in the children's everyday environments. Kasari et al. noted that language samples are highly valid for use with children with ASD, and provide useful measures of expressive language in both verbal and minimally verbal children. Indeed, as Heilmann, Miller, and Nockerts [24] claimed that "language sampling and analysis has unparalleled validity for measuring language use" and has long been considered the gold standard approach for evaluating spoken language skills in children [25-27]. Given the pressing need to support, measure, and report the development of spoken language in children with ASD receiving comprehensive intervention programs, the aim of this review was to

document (a) the skills targeted and (b) assessment tools used to measure spoken language outcomes for children with ASD receiving comprehensive intervention programs in controlled studies.

Method

We conducted a systematic review of the literature according to the methods outlined by Pickering and Byrne [28].

Inclusion Criteria

The inclusion criteria were: (a) articles published between 1987 (year in which Lovaas published original article on comprehensive applied behavior analysis intervention for children with ASD [29]) to July 2015 in peer-reviewed journals, (b) the study included comprehensive intervention (i.e. addressing core difficulties in autism including language, social, cognition and play; [30]) based on the principals of applied behaviour analysis (e.g. EIBI, discrete trial training), (c) the study contained a control group, (d) participants had a diagnosis of ASD (i.e. autism, autistic disorder, PDD-NOS) and (d) the majority of participants commenced intervention prior to formal schooling.

Search Procedures

We completed searches of Medline, Psycinfo, and ERIC using the search terms “autis*” AND “child*” AND [“intensive behav* intervention” OR “intensive” OR “discrete trial training” and “applied behave* analysis”]. A total of 1077 articles were returned from initial searches. Article titles and abstracts from databases were reviewed and 82 articles were short-listed for further analysis to determine eligibility against inclusion criteria. Inter-rater reliability was calculated by comparing the papers for the Medline database independently selected by the first and second authors for short-listing, yielding agreement of 100%. From the 82 short-listed

articles, 32 were found to be reviews, meta-analyses, or more general discussions of behavioural intervention practices; 27 articles were deemed not to meet inclusion criteria; and 23 articles were progressed for data extraction. Reasons for exclusion included: (a) the study did not contain a control group, (b) the intervention was not based on principles of applied behaviour analysis, (c) the study was continuing longitudinal and therefore participants were beyond the desired age range or the outcome measures had already been included from another study, and (d) the study did not include child-outcome variables. We also completed a hand search of the reference lists of review articles identified during the search and those of included studies, yielding 5 additional articles that met the inclusion criteria.

Data Extraction

The following information was extracted for each study: (a) number of participants and participants' ages and diagnoses; (b) treatment type, frequency, duration and personnel responsible for implementation; (c) comparison intervention/s; (d) dependent variables targeted; and (e) assessment tools used to measure dependent variables.

Results

As presented in Table 1, 504 children with ASD were included in treatment groups across the 23 studies. Substantial variability in methods used to confirm diagnosis is evident, ranging from inclusion based on community-based diagnosis of ASD to community diagnosis plus confirmation as part of the study using ASD-specific diagnostic tools. Although the breakdown of sub-groups (where possible under DSM-IV [31]) was provided for a small number of studies (e.g., Dawson et al. [5]), in most cases readers are reliant on assessment results at Time 1 to infer the children's abilities. The children received a range of comprehensive interventions delivered by trained clinicians, volunteers, and/or parents based on clinician engagement of between 1-40 hours per week, over a period of between 12 weeks to approximately 3 years. The two most

common comparison interventions were treatment as usual (7 studies) and eclectic interventions (7 studies), followed by programs delivered in special education settings (5 studies).

<insert table 1 about here>

Dependent Variables Assessed

The dependent variables (DVs) targeted in each study are listed in Table 1, using the terms reported in each study. To summarise the variables, we categorised them according to the skills targeted, resulting in the seven categories presented in Figure 1. Social-communication skills, intellectual functioning, and daily living skills were the three most frequently targeted aspects of development. The social-communication category included broad measures of ‘communication’ (42 DVs), ‘expressive’ and or ‘receptive’ language (20 DVs), following instructions (6 DVs), object labelling (6 DVs), vocabulary (8 DVs) requesting (2 DVs), response to name (2 DVs), and ‘words’ (2 DVs).

< insert figure 1 about here>

Assessment tools utilised

As illustrated in Figure 2 the five most frequently used assessment tools included measures of intellectual functioning and receptive and expressive language, as well as a set of customised assessment tools designed to measure a range of adaptive and maladaptive behaviours. Note that only assessment tools that were used to assess dependent variables in each study are presented in Table 1 and Figure 2, not tools used for the sole purpose of participant characterisation. Although assessments of intellectual functioning (e.g., Mullen Scales of Early Learning [32]) and adaptive

behaviour (e.g., Vineland Adaptive Behavior Scales [15]) include expressive and receptive language domains, few dedicated language assessments were included. Of the assessment tools designed specifically to measure language, the Reynell Developmental Language Scales [norm-referenced clinician-administered receptive and expressive language scales; 33] was the most commonly used (9 studies), followed by the MacArthur Bates – Communicative Development Inventory (norm-referenced parent-report tool assessing expressive and receptive vocabulary; 2 studies), Peabody Picture Vocabulary Test (norm-referenced receptive vocabulary assessment; 2 studies), the Clinical Evaluation of Language Fundamentals (norm-referenced clinician-administered broad spectrum receptive and expressive language assessment ; 1 study), and the Receptive-Expressive Emergent Language Scale (norm-referenced parent-interview receptive and expressive language sub-tests; 1 study). Four customised assessment tools (e.g., Early Learning Measure) included specific measures of language ability (e.g., number of recognisable words, verbal imitation, expressive object labelling). However, none of the studies in this review included the collection and analysis of language samples in natural environments.

<insert figure 2 about here>

Discussion

Our aim was to document the intervention targets and assessment tools used to measure spoken language outcomes for children with ASD receiving comprehensive intervention programs. Our systematic review, spanning the 28 years since Lovaas' original study of applied behavior analysis [29] for children with ASD, yielded 23 studies that met our inclusion criteria. The results indicate that although supporting social-communication development is consistently a primary target of intervention, the dependent variables selected and the assessment tools used to measure progress focus on broad measures of communication development, predominantly through the use of standardised assessments in clinical settings and parent report. No studies used language sampling and analysis, despite its validity for measuring language use in everyday

settings [24], thus limiting our capacity to draw comprehensive conclusions regarding the extent to which participants achieved the ultimate communication goal of spontaneous, flexible, directed, and effective spoken language for a range of purposes across all life contexts.

Communication as a dependent variable

In most studies, children's development of spoken language was assessed under a broad dependent variable such as 'communication' or 'expressive language.' This approach is understandable, given the statistical imperative to preserve power by minimising the number of dependent variables in trials featuring relatively small numbers of participants, as is common in ASD research. This issue is particularly pertinent in evaluations of comprehensive intervention programs targeting multiple developmental domains, with each requiring measurement. Nevertheless, there is a risk that in selecting a broad dependent variable, important information regarding children's functional skills across the domains of morphology, semantics, syntax, phonology, and pragmatics will be lost or masked. The alternative is to add to these broader dependent variables with more specific and descriptive language targets. Lovaas [29], for example, included a video-recorded play-based observation measure of "recognizable words as part of the pre-treatment assessment battery, defined to include any recognizable word, independent of whether the subject used it in a meaningful context or for communicative purposes" (pp.50). Although the operationalisation of the dependent variable is at odds with the definition for spoken language used in this review (spontaneous, flexible, directed, etc), its inclusion demonstrates the potential value of including descriptive, easily interpreted, variables in intervention research. Notably, a brief scan of Table 1 reveals that earlier studies more often included dependent variables related to specific communicative behaviours (e.g., vocal imitation, expressive object labelling, use of words) than later studies. Given the increasing awareness of the need to study individual differences in intervention outcomes, and the calls for the development of

more sensitive and socially valid outcome measures [11], it may be timely to consider what can be learned from these earlier endeavours to document intervention outcomes.

Assessing communication outcomes

The majority of studies used standardised assessments to measure intervention outcomes. These assessment tools have clear benefits including standardised administration procedures, the capacity to compare children's functioning with peers of similar age, and the opportunity to compare the findings across different intervention studies [13]. The Reynell Developmental Language Scales [33], for example, was the third most commonly used assessment tool across the studies reviewed. It is a norm-referenced assessment, appropriate for young children (2 to 7 years, 6 months), which includes a mix of play-based activities using a range of stimulus activities designed to assess both receptive and expressive language skills, including understanding and production of selected vocabulary and grammatical features. Kjellmer et al., [34] suggested that the use of concrete objects and pictures may be appealing to children with neurodevelopmental disorders. However, there are documented limitations with the use of standardised assessments, including the contexts in which behaviours are assessed and the potential confound of test-taking ability on test performance, which may limit their capacity to capture each child's functional capacity across a range of contexts.

Several authors have advocated for greater use of language sampling and analysis in ASD research, but this approach did not feature in any of the studies reviewed. Considering our ultimate goal for children with ASD to become successful communicators in everyday contexts, the importance of spontaneous language sampling and analysis cannot be ignored (see also [35]). Condouris, Meyer, and Tager-Flusberg [13] raised concerns that language sampling and analysis is generally time consuming and presented evidence that standardised language assessments may correlate well with measures derived from language samples. However, the finding was based on

a study involving children who were all able to complete the selected standardised assessments within their age level, rather than the broader spectrum of children with ASD, including those with intellectual disability or those who show limited verbal skills. Results from a study of minimally verbal children with ASD [36] showed that these children produced more spontaneous communication than elicited communication, indicating that standardised tests may at times underestimate the children's expressive language abilities. Furthermore, results from standardised tests will fail to yield the descriptive detail needed for careful, individualised intervention planning or progress monitoring [37]. Technology has the potential to assist with both the collection (e.g., Language Environment Analysis; LENA) and analysis (e.g., Systematic Analysis of Language Transcripts; SALT, [38]) of language samples and thus help to reduce the time demands. Therefore, language sampling need not be considered an alternative to standardised assessments, but rather a crucial complementary method designed to provide a comprehensive picture of children's intervention outcomes. Language sampling and analysis can help to confirm the results of standardised assessments, provide rich descriptive information for intervention planning, and contribute to detailed monitoring of children's response to intervention [27, 58]. Furthermore, evaluating children's language in everyday interactions with parents, teachers, and peers would also help to address McConachie et al's., [11] call for the use of ecologically valid outcome measures in ASD research.

Conclusion

Given that promoting spoken language development in children with ASD is a primary goal of comprehensive interventions, it is imperative that the way it is targeted (i.e., in terms of dependent variables) and measured (i.e., in terms of assessment tools) is the subject of careful consideration when critically appraising the results of studies to date, and designing future studies. The results of this review indicate that there is an opportunity to move the field forward by supplementing the use of standardised assessments with language sampling and analysis in

children's everyday environments. Such an approach would afford parents, educators, and researchers more accurate and detailed insights into the real-world impacts of comprehensive interventions for children with ASD.

References

1. Paul, R. and C.F. Norbury, Language disorders from infancy through adolescence. Listening, speaking, reading, writing, and communicating (4th Ed). Baltimore: Mosby Elsevier; 2012.
2. Howlin, P., et al., *Adult outcome for children with autism*. Journal of Child Psychology and Psychiatry. 2004;45:212-229.
3. **National Autism Center, National Standards Report. Randolph, MA: National Autism Centre; 2015. *This report presents a comprehensive review of the evidence-base for ASD interventions based on research conducted in the period 2007-2012*.
4. Odom SL, et al. Evidence-Based Practices in Interventions for Children and Youth with Autism Spectrum Disorders. Preventing School Failure: Alternative Education for Children and Youth. 2010;54(4):275-282.
5. Dawson G, et al. Randomized, controlled trial of an intervention for toddlers with autism: the Early Start Denver Model. Pediatrics. 2010;125(1):e17-23.
6. Bondy A, Frost L. The Picture Exchange Communication System. Behav Modif. 2001;25(5):725-44.
7. Maglione MA, et al. Nonmedical interventions for children with ASD: recommended guidelines and further research needs. Pediatrics. 2012;130 Suppl 2:S169-78.
8. Reichow B, et al. Early intensive behavioral intervention (EIBI) for young children with autism spectrum disorders (ASD). Cochrane Database Syst Rev. 2012;10: CD009260.
9. Tonge BJ, et al. A review of evidence-based early intervention for behavioural problems in children with autism spectrum disorder: the core components of effective programs, child-

focused interventions and comprehensive treatment models. *Curr Opin Psychiatry*. 2014;27(2):158-65.

10. Trembath, D. and G. Vivanti, *Problematic but predictive: individual differences in children with autism spectrum disorders*. *International Journal of Speech Language Pathology*. 2014;16(1):57-60.
11. **McConachie H, et al. Systematic review of tools to measure outcomes for young children with autism spectrum disorder. *Health Technol Assess*. 2015;19(41): 1-506. *The most comprehensive review of assessment tools for individuals with ASD currently available*.
12. Matson, J.L. and R.D. Rieske, *Are outcome measures for early intensive treatment of autism improving?* *Research in Autism Spectrum Disorders*, 2014. 8(3): p. 178-185.
13. Condouris K, Meyer E, Tager-Flusberg H. The relationship between standardized measures of language and measures of spontaneous speech in children with autism. *American Journal of Speech-Language Pathology*. 2003;12(3):349-358.
14. Matson, J.L., *Determining treatment outcome in early intervention programs for autism spectrum disorders: A critical analysis of measurement issues in learning based interventions*. *Research in Developmental Disabilities*, 2007. 28(2): p. 207-218.
15. Sparrow SS, Cicchetti DV, Balla DA. *Vineland Adaptive Behavior Scales - 2nd Edition (Vineland - II)*. Livonia, MN: Pearson Assessments; 2005.
16. Neisworth JT, Bagnato SJ. The mismeasure of young children - The authentic assessment alternative. *Infants and Young Children*. 2004;17(3):198-212.
17. Fenson L, et al. *Technical Manual for the MacArthur communicative development inventories*. San Diego: San Diego State University; 1991.
18. Green J, et al. Parent-mediated communication-focused treatment in children with autism (PACT): a randomised controlled trial. *Lancet*. 2010;375(9732):2152-60.

19. McConachie H, et al. A controlled trial of a training course for parents of children with suspected autism spectrum disorder. *J Pediatr*. 2005;147(3):335-40.
20. Filipek PA, et al. The screening and diagnosis of autistic spectrum disorders. *Journal of Autism and Developmental Disorders*. 1999;29(6):439-484.
21. Raising Children Network. Parent guide to therapies. 25/05/15]; Available from: http://raisingchildren.net.au/parents_guide_to_therapies/parents_guide_to_therapies.html
22. NIASA, N.I.f.A.S.a.A., National Autism Plan for Children (NAPC): Plan for the identification, assessment, diagnosis and access to early interventions for pre-school and primary school aged children with autism spectrum disorders (ASD). London: The National Autistic Society; 2003.
23. Kasari C, et al. Assessing the Minimally Verbal School-Aged Child With Autism Spectrum Disorder. *Autism Research*. 2013;6(6):479-493.
24. Heilmann, J.J., J.F. Miller, and A. Nockerts, *Using language sample databases*. *Lang Speech Hear Serv Sch*. 2010; **41**(1):84-95.
25. Tager-Flusberg, H. and J. Cooper, *Present and future possibilities for defining a phenotype for specific language impairment*. *J Speech Lang Hear Res*. 1999;**42**(5):1275-8.
26. Dunn, M., et al., *The use of spontaneous language measures as criteria for identifying children with specific language impairment: an attempt to reconcile clinical and research incongruence*. *J Speech Hear Res*. 1996;**39**(3):643-54.
27. Miller, J.F., *Progress in assessing, describing, and defining child language disorder*, in *Assessment of communication and language* K.N. Cole, P.S. Dale, and D.J. Thal, Editors. Baltimore. 1996 p. 309-324.
28. Pickering CM, Byrne J. The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early career researchers. *Higher Education Research and Development*. 2014;33:534-548.

29. Lovaas OI. Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology*. 1987;55(1):3-9.
30. Rogers SJ, Vismara LA. Evidence-based comprehensive treatments for early autism. *Journal of Clinical Child & Adolescent Psychology*. 2008;37(1):8-38.
31. American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC: American Psychiatric Association; 1994.
32. Birnbrauer JS, Leach DJ. The Murdoch Early Intervention Project after 2 Years. *Behaviour Change*. 1993;10(2):63-74.
33. Smith T, Groen AD, Wynn JW. Randomized Trial of Intensive Early Intervention for Children with Pervasive Developmental Disorder. *American Journal on Mental Retardation*. 2000;105(4):269-285.
34. Eikeseth S, et al. Intensive behavioral treatment at school for 4- to 7-year-old children with autism. A 1-year comparison controlled study. *Behav Modif*. 2002;26(1):49-68.
35. Sallows GO, Graupner TD. Intensive behavioral treatment for children with autism: Four-year outcome and predictors. *American Journal of Mental Retardation*. 2005;110(6):417-438.
36. Howard JS, et al. A comparison of intensive behavior analytic and eclectic treatments for young children with autism. *Research in Developmental Disabilities*. 2005;26:359-383.
37. Cohen H, Amerine-Dickens M, Smith T. Early intensive behavioral treatment: Replication of the UCLA model in a community setting. *Journal of Developmental and Behavioral Pediatrics*. 2006;27(2):S145-S155.
38. Eldevik S, et al. Effects of low-intensity behavioral treatment for children with autism and mental retardation. *J Autism Dev Disord*. 2006;36(2):211-24.

39. Eikeseth S, et al. Outcome for children with autism who began intensive behavioral treatment between ages 4 and 7: a comparison controlled study. *Behav Modif.* 2007;31(3):264-78.
40. Zachor DA, et al. Change in autism core symptoms with intervention. *Research in Autism Spectrum Disorders.* 2007;1(4):304-317.
41. Reed P, Osborne LA, Corness M. Brief report: Relative effectiveness of different home-based behavioral approaches to early teaching intervention. *Journal of Autism and Developmental Disorders.* 2007;37(9):1815-1821.
42. Reed P, Osborne LA, Corness M. The real-world effectiveness of early teaching interventions for children with autism spectrum disorder. *Exceptional Children.* 2007;73(4):417-433.
43. Magiati I, Charman T, Howlin P. A two-year prospective follow-up study of community-based early intensive behavioural intervention and specialist nursery provision for children with autism spectrum disorders. *J Child Psychol Psychiatry.* 2007;48(8):803-12.
44. Remington B, et al. Early intensive behavioral intervention: outcomes for children with autism and their parents after two years. *Am J Ment Retard.* 2007;112(6):418-38.
45. Fava L, et al. 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.* 2011;5(4):1479-1492.
46. Hayward D, et al. Assessing progress during treatment for young children with autism receiving intensive behavioural interventions. *Autism.* 2009;13(6):613-33.
47. Peters-Scheffer N, et al. Low intensity behavioral treatment supplementing preschool services for young children with autism spectrum disorders and severe to mild intellectual disability. *Research in Developmental Disabilities.* 2010;31(6):1678-1684.

48. Zachor DA, Itzhak EB. Treatment approach, autism severity and intervention outcomes in young children. *Research in Autism Spectrum Disorders*. 2010;4(3):425-432.
49. Eldevik S, et al. Outcomes of behavioral intervention for children with autism in mainstream pre-school settings. *J Autism Dev Disord*. 2012;42(2):210-20.
50. Reed P, Osborne L. Impact of severity of autism and intervention time-input on child outcomes: comparison across several early interventions. *British Journal of Special Education*. 2012;39(3):130-136.
51. Rogers SJ, et al. Effects of a Brief Early Start Denver Model (ESDM)-Based Parent Intervention on Toddlers at Risk for Autism Spectrum Disorders: A Randomized Controlled Trial. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2012;51(10):1052-1065.
52. Reitzel J, et al. Pilot randomized controlled trial of a Functional Behavior Skills Training program for young children with Autism Spectrum Disorder who have significant early learning skill impairments and their families. *Research in Autism Spectrum Disorders*. 2013;7(11):1418-1432.
53. Mullen, E., *Mullen Scales of Early Learning* 1995, Circle Pines, MN: AGS.
54. Reynell, J. and M. Huntley, *Reynell developmental language scales - Revised*. 3rd ed 1987, Windsor: NFER-Nelson.
55. Kjellmer L, et al. Language comprehension in preschoolers with autism spectrum disorders without intellectual disability: Use of the Reynell Developmental Language Scales. *Research in Autism Spectrum Disorders*. 2012;6(3):1119-1125.
56. Kim, S.H., D. Junker, and C. Lord, Observation of Spontaneous Expressive Language (OSEL): a new measure for spontaneous and expressive language of children with autism spectrum disorders and other communication disorders. *J Autism Dev Disord*. 2014;44(12):3230-44.

57. Chiang, H.M., Naturalistic observations of elicited expressive communication of children with autism: an analysis of teacher instructions. *Autism*. 2009;13(2):65-78.
58. Westerveld, M., Spontaneous language sampling and analysis in children: From research to practice. *ACQuiring Knowledge in Speech, Language and Hearing*. 2011;13(2):63-67.
59. Miller, J.F., et al., Assessing Language Production using SALT Software. A Clinician's Guide to Language Sample Analysis. New Zealand version. Middleton, WI: SALT Software; 2012.

Table 1. Summary of Dependent Variables and Outcome Measures used in evaluation of comprehensive interventions

Study	Design	ASD Treatment Group ^{a#}	Treatment	Comparison Treatment	Dependent Variable/s	Outcome Measure
Lovaas, 1987 [29]	Non-RCT	n = 19 Age: mean 34.6 months Diagnosis: Autism (DSM-III)	Applied Behaviour Analysis delivered 1:1 by trained therapists. Approximately 40 hours per week for 2 or more years.	Low intensity Applied Behaviour Analysis	Intellectual Functioning	Wechsler Intelligence Scale for Children-Revised Stanford-Binet Intelligence Scale Merrill-Palmer Pre-School Performance Test Wechsler Pre-School Scale Vineland Social Maturity Scale Gesell Infant Development Scale Bayley Scales of Infant Development Cattell Infant Intelligence Scale Peabody Picture Vocabulary Test Parent report of educational placement Vineland Adaptive Behavior Scales
Birnbrauer & Leach, 1993 [39]	Non-RCT	n = 9 Age: 39 months (range 32-47) Diagnosis: Autistic disorder or PDD-NOS (DSM-III-R)	Murdoch Early Intervention Program, delivered 1:1 by parents and volunteers. Mean 18.72 hours per week for 24 months.	Treatment as usual	Educational placement Communication, daily living, social, and motor skills composite score Intellectual functioning	Bayley Scales of Infant Development Stanford Binet Intelligence Scale Leiter International Performance Scale Peabody Picture Vocabulary Test Wechsler Preschool and Primary Scale of Intelligence Wechsler Intelligence Scale for Children Reynell Developmental Language Scale Receptive-Expressive Emergent Language Scale Personality Inventory for Children
Smith, Groen, & Wynn, 2000 [40]	RCT	n = 15 Age: 36.07(6.0) Diagnosis: Autism/PDD-NOS	UCLA early intensive intervention, delivered 1:1 by clinicians. Mean 24.52 hours per week in the first the 12 months, with a reduction in hours thereafter.	Parent administered UCLA early intensive intervention	Receptive and expressive language Emotional, behavioral, social, and cognitive adjustment Play, self-stimulatory behaviour, withdrawal, compliance, and imitation under controlled conditions ASD symptoms Intellectual functioning	Customised video assessment involving structured play interaction with clinician. Customised rating scale based on DSM-III-R criteria Bayley Scales of Infant Development Stanford Binet Intelligence Scale Merrill-Palmer Scale of Mental Tests Reynell Developmental Language Scales Vineland Adaptive Behavior Scales
Eikeseth, Smith, Jahr, & Eldevik, 2002 [41]	Non-RCT	n = 13 Age: 66.31(11.31) Diagnosis: Childhood autism (ICD-10)	Behavioral intervention delivered 1:1 by teachers, aides, and parents. Mean 28.52 hours per week for 12 months.	Eclectic intervention	Receptive and expressive language Communication, daily living, social, and motor skills domain and composite scores Verbal imitation, nonverbal imitation, following verbal instructions, and expressive object labelling	Early Learning Measure
Sallows & Graupner, 2005 [42]	RCT	n = 13 Age: 33.23(3.89) Diagnosis: ASD (DSM-IV)	EIBI delivered 1:1 by clinicians in clinical setting. Mean 38.6 hours per week for 2 years.	EIBI delivered by trained parents.	Intellectual functioning	Wechsler Preschool and Primary Scale of Intelligence Wechsler Intelligence Scale for Children -Revised Bayley Scales of Infant Development Merrill-Palmer Pre-School Performance Test Reynell Developmental Language Scale Vineland Adaptive Behavior Scales
					Receptive and expressive language	Bayley Scales of Infant Development Wechsler Preschool and Primary Scale of Intelligence Wechsler Intelligence Scale for Children (3 rd Ed.) Bayley Scales of Infant Development (2 nd Ed.) Reynell Developmental Language Scales

					Communication, daily living, and social skills domain and composite scores Autism symptoms Verbal imitation, nonverbal imitation, following verbal instructions, and expressive object labelling Intellectual functioning	Clinical Evaluation of Language Fundamentals (3 rd Ed.) Vineland Adaptive Behavior Scales Autism Diagnostic Interview - Revised Early Learning Measure
Howard, Sparkman, Cohen, Green, & Stanislaw, 2005 [43]	Non-RCT	n = 29 Age: 30.86(5.16) Diagnosis: ASD (DSM-IV)	Intensive behaviour analytic treatment, delivered 1:1 delivered by clinicians. 25-40 per week for approximately 12 months.	Public school classrooms for children with ASD. Special education classrooms for children with a range of disabilities.	Receptive and expressive language Communication, daily living, social, and motor skills domain and composite scores	Bayley Scales of Infant Development Wechsler Preschool and Primary Scale of Intelligence Differential Abilities Scale Stanford Binet Intelligence Scale (4 th Ed.) Merrill-Palmer Scale of Mental Tests Reynell Developmental Language Scales Vineland Adaptive Behavior Scales
Cohen, Amerine-Dickens, & Smith, 2006 [44]	Non-RCT	n = 21 Age: 30.2(5.8) Diagnosis: Autistic disorder or PDD-NOS	EIBI delivered 1:1 by clinicians. 35-40 hours per week for 3 or more years.	Special education classes at local schools.	Intellectual functioning Visual-spatial skills Expressive and receptive language Communication, daily living, and social skills domain and composite scores Intellectual functioning	Bayley Scales of Infant Development Wechsler Preschool and Primary Scale of Intelligence – Revised Merrill-Palmer Scale of Mental Tests Reynell Developmental Language Scales Vineland Adaptive Behavior Scales
Eldevik, Eikeseth, Jahr, & Smith, 2006 [45]	Non-RCT	n = 13 Age: 53(9.5) Diagnosis: Childhood autism (ICD-10)	Behavioural intervention delivered 1:1 by clinicians. Mean 15 hours per week for 2 years.	Eclectic intervention	 Receptive and expressive language Adaptive behaviour: words, affection, toy play, peer play, stereotypical behaviours, severe tantrums, toilet training "Degree of mental retardation" Communication, daily living, social skill, and motor skills composite score Autism symptoms	Bayley Scales of Infant Development Stanford-Binet Intelligence Scale Wechsler Preschool and Primary Scale of Intelligence Stanford-Binet Intelligence Scale Wechsler Intelligence Scale for Children-Revised Merrill-Palmer Pre-School Performance Test Reynell Developmental Language Scale Psycho Educational Profile-Revised Customised questionnaire and scoring system based on archival data, parent report, or direct observation ICD-10 classifications for 'mental retardation' Vineland-II Childhood Autism Rating Scale
Eikeseth, Smith, Jahr, & Eldevik, 2007 [46]	Non-RCT	n = 35 Age: 4 years; 5 months (range 2;0 – 7;4) Diagnosis: Autism (ICD-10)	Applied Behavior Analysis delivered by school staff. Approximately 23 hours per week for 1 year.	Treatment as usual		
Zachor, Ben-Itzhak, Rabinovich, & Lahat, 2007 [47]	Non-RCT	n = 45 Age: 25.1(3.9) Diagnosis: Autism (DSM-IV)	Applied Behavior Analysis delivered 1:1 by clinicians. Approximately 35 hours per week for 12 months.	Eclectic intervention	Autism symptoms Intellectual functioning	Autism Diagnostic Observation Schedule Bayley Scales of Infant Development Stanford Binet Intelligence Scale (4 th Ed.)
Reed, Osborne, & Corness,	Non-RCT	n = 12 Age: 40 months (range 32-47)	Applied Behavior Analysis delivered 1:1 by trained tutors. Mean 30.4 hours per	Special nursery, Portage program	Autism symptoms Imitation, Perception, Fine and Gross Motor Skills, Eye-Hand Coordination and Nonverbal and Verbal	Gilliam Autism Rating Scale Psycho-Educational Profile

2007 [48]		Diagnosis: ASD	week for 24 months.		Conceptual Ability domain and composite scores) Intellectual functioning (Verbal Comprehension, Early Number Concepts, Picture Matching and Naming Vocabulary Sub-tests) Communication, daily living, social, and motor skills domain and composite scores	British Abilities Scale (2 nd Ed.) Early Years Battery Vineland Adaptive Behavior Scales
Reed, Osborne, & Corness, 2007 [49]	Non-RCT	n = 14 Age: 42.9(14.8) Diagnosis: ASD (community diagnosis)	‘High intensity’ Applied Behavior Analysis delivered 1:1 by trained tutors. Approximately 20-40 hours per week for 9-10 months.	‘low intensity’ Applied Behavior Analysis	Educational achievement Autism symptoms Imitation, Perception, Fine and Gross Motor Skills, Eye-Hand Coordination and Nonverbal and Verbal Conceptual Ability domain and composite scores) Intellectual functioning (Verbal Comprehension, Early Number Concepts, Picture Matching and Naming Vocabulary Sub-tests) Communication, daily living, social, and motor skills domain and composite scores	British Abilities Scale (2 nd Ed.) Early Years Battery Gilliam Autism Rating Scale Psycho-Educational Profile British Abilities Scale (2 nd Ed.) Ed Early Years Battery Vineland Adaptive Behavior Scales
Magiati, Charman, & Howlin, 2007 [50]	Prospective follow-up study	n = 25 Age: 38(7.2) Diagnosis: Autism (community diagnosis)	EIBI delivered 1:1 by clinicians. Mean 32.4 hours per week for 24 months.	Eclectic intervention	Autism symptoms Intellectual functioning Communication, daily living, social, and motor skills domains and composite scores Receptive vocabulary Expressive vocabulary Play skills	Autism Diagnostic Interview - Revised Bayley Scales of Infant Development Wechsler Preschool and Primary Scale of Intelligence – Revised Merrill-Palmer Scale of Mental Tests Vineland Adaptive Behavior Scales British Picture Vocabulary Scale-II Expressive One-Word Picture Vocabulary Test-Revised Symbolic Play Test – II Test of Pretend Play Stanford Binet Intelligence Scale (4 th Ed.) Reynell Developmental Language Scales (3 rd Ed.) Vineland Adaptive Behaviour Scales
Remington et al., 2007 [51]	Non-RCT	n = 23 Age: 35.7(4.0) Diagnosis: Autism, based on community diagnosis and ADI-R.	Early Intensive Behavioral Intervention delivered 1:1 by trained tutors and parents. Mean 25.6 hours per week for 2 years.	Treatment as usual	Intellectual functioning Expressive and receptive language Communication, social, and daily living skills domain and composite scores Behaviour Autism symptoms Joint attention (initiation and response)	Nisonger Child Behaviour Rating Form Developmental Behaviour Checklist Autism Screening Questionnaire Early Social Communication Scales Autism Diagnostic Observation Schedule Griffith Mental Developmental Scales for ages 2 - 8 MacArthur Bates – Communicative Development Inventories Customised video-based challenging behaviours assessment Vineland Adaptive Behavior Scales
Fava et al., 2011 [52]	Non-RCT	n = 12 Age: 52(19.5) Diagnosis: Autism or PDD-NOS	EIBI delivered 1:1 and small group across settings by staff and parents. Approximately 26 hours per week for 12 months.	Eclectic intervention	Autism symptoms Intellectual functioning Receptive and expressive language Frequency and function of challenging behaviours Communication, daily living, social, and motor skills domain and composite scores Psychopathological comorbidities Parenting stress	Autism Screening Questionnaire Early Social Communication Scales Autism Diagnostic Observation Schedule Griffith Mental Developmental Scales for ages 2 - 8 MacArthur Bates – Communicative Development Inventories Customised video-based challenging behaviours assessment Vineland Adaptive Behavior Scales Child Behaviour Checklist Parenting Stress Index – Short Form Bayley Scales of Infant Development Wechsler Preschool and Primary Scale of Intelligence – Revised Merrill-Palmer Scale of Mental Tests Reynell Developmental Language Scales Vineland Adaptive Behavior Scales
Hayward, Eikeseth, Gale, & Morgan, 2009 [53]	Non-concurrent multiple baselin	n = 23 Age: 35.7(6.2) Diagnosis: Autism (ICD-10)	UCLA Applied Behavior Analysis delivered 1:1 by clinicians. Approximately 36 hours per week for 12 months.	UCLA Applied Behavior Analysis delivered by parent-	Intellectual functioning Receptive and expressive language Communication, daily living, social, and motor skills	Bayley Scales of Infant Development Wechsler Preschool and Primary Scale of Intelligence – Revised Merrill-Palmer Scale of Mental Tests Reynell Developmental Language Scales Vineland Adaptive Behavior Scales

e				organised clinicians	domain and composite scores 'non-vocal imitation', 'receptive instruction', 'vocal imitation' & 'expressive labels'	Session treatment data
Dawson et al., 2010 [5]	RCT	n = 24 Age: 23.1(3.9) Diagnosis: 21 ASD, 3 PDD-NOS (DSM-IV)	Early Start Denver Model delivered 1:1 by trained clinicians and parents. Mean 15.2 clinician hours per week for 24 months.	Treatment as usual	Autism symptoms Intellectual functioning Communication, daily living, social skill, and motor skills composite score Parent questionnaire on repetitive behaviours	Autism Diagnostic Observation Scale Mullen Scales of Early Learning Vineland-II
Peters-Scheffer, Didden, Mulders, & Korzilius, 2010 [54]	Non-RCT	n = 12 Age: 53.5(5.52) Diagnosis: ASD (DSM-IV)	Discrete trial training, delivered 1:1 by supervised preschool staff. Mean 6.29 hours per week for 8 months, in addition to regular eclectic early intervention program.	Regular eclectic early intervention program.	Intellectual Functioning Communication, daily living, and social skills domain and composite scores Emotional and behaviour problems Communication, social behaviour, and stereotyped behaviour	Repetitive Behavior Scale Bayley Scales of Infant Development (Dutch version) SON 2.5-7 Vineland Adaptive Behavior Scales (Dutch version) Child Behaviour Checklist (Dutch version) Scale of Pervasive Developmental Disorder in Mentally Retarded Persons (Dutch version) Autism Diagnostic Observation Schedule Vineland Adaptive Behavior Scales
Zachor & Ben Itzhak, 2010 [55]	Non-RCT	n = 45 Age: 25.1(3.9) Diagnosis: Autism (DSM-IV)	Applied Behavior Analysis delivered 1:1 by therapists with certificate in ABA. Approximately 20 hours per week for 12 months.	Eclectic intervention	Autism symptoms Communication, daily living, social skill, and motor skills Visual reception, fine motor, receptive language, and expressive language skills	Mullen Scales of Early Learning
Eldevik, Hastings, Jahr., & Hughes, 2012 [56]	Non-RCT	n = 31 Age: 42.2(9.0) Diagnosis: Autism or PDD-NOS (ICD-10)	Early Intensive Behavioral Intervention delivered 1:1 by school staff trained by psychologist. Mean 13.6 hours of treatment per week for 2 years.	Treatment as usual	Intellectual functioning Communication, daily living, and social skills domains and composite	Bayley Scales of Infant Development (2 nd or 3 rd Edition) Stanford Binet Intelligence Scales Wechsler Preschool and Primary Scale Intelligence-Revised Vineland Adaptive Behavior Scales or Vineland-II
Reed & Osborne., 2012 [57]	Non-RCT	n = 14 Age = 39(6.9) Diagnosis: Autistic disorder or PDD-NOS (DSM-IV)	Applied Behaviour Analysis, delivered 1:1 by trained tutors. Approximately 30 hours per week for 9 months.	Special nursery Program, Portage Program, Local authority approach	Cognitive development Educational achievement Communication, daily living, social skill, and motor skills composite score	Psycho-Educational Profile British Abilities Scale (2 nd Ed.) Early Years Battery Vineland-II
Rogers et al., 2012 [58]	RCT	n = 49 Age: 12-24 Diagnosis: 'At risk for ASD'	Early Start Denver Model Based parent intervention. Parents coached to deliver treatment during 12 x 1 hour training sessions over 12 weeks.	Treatment as usual	Social affect, restricted and repetitive behaviours Intellectual functioning Expressive and receptive vocabulary Communication, daily living, social, and motor skills	Autism Diagnostic Observation Schedule Mullen Scales of Early Learning Macarthur Bates Communicative Development Inventory Vineland-II
Reitzel et al., 2013 [59]	RCT	n = 8 Age: 57.9(13.1) Diagnosis: Autistic Disorder	Functional Behavioural Skills Training delivered by a teacher, a parent coach and child prompters. Weekly 2hr group sessions for 4 months.	Treatment as usual	Communication, daily living, social, and motor skills Severity of problem behaviours Adaptive behaviour (requesting an item or activity, hand washing, eating at table, response to name, toileting)	Vineland-II Developmental Behaviour Checklist Customised Functional Behaviour Skills Assessment

^o Age in months (mean/SD) unless otherwise indicated. ^φ Table presents diagnostic terminology used in each article.

Figure 1. Summary of dependent variables targeted across the 23 studies.

Figure 2. Assessment tools used to measure treatment outcomes across studies.

Notes: (a) Assessment tools published in multiple versions/editions have been combined to yield frequency count, (b) 'custom' refers to outcome measures developed for use in the study, rather than published assessment tools, (c) 'other' refers to assessment tools that were used in fewer than 3 studies included in the review, (d) standardised tests, including original and revised editions where relevant: *Wechsler Intelligence Scale for Children*, *Wechsler Preschool and Primary Scale of Intelligence*, *Wechsler Pre-School Scale*; *Bayley Scales of Infant Development*; *Reynell Developmental Language Scales*; *Stanford-Binet Intelligence Scale*; *Autism Diagnostic Observation Schedule*; *Merrill-Palmer Scale of Mental Tests* *British Abilities Scale*; *Psycho-Educational Profile*; *Merrill-Palmer Pre-School Performance Test*; *Mullen Scales of Early Learning*.