



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Interventions based on the Theory of Mind cognitive model for autism spectrum disorder (ASD)

Citation for published version:

Fletcher-Watson, S, McConnell, F, Manola, E & McConachie, H 2014, 'Interventions based on the Theory of Mind cognitive model for autism spectrum disorder (ASD)', *Cochrane Database of Systematic Reviews*.
<https://doi.org/10.1002/14651858.CD008785.pub2>

Digital Object Identifier (DOI):

[10.1002/14651858.CD008785.pub2](https://doi.org/10.1002/14651858.CD008785.pub2)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Cochrane Database of Systematic Reviews

Publisher Rights Statement:

© Fletcher-Watson, S., McConnell, F., Manola, E., & McConachie, H. (2014). Interventions based on the Theory of Mind cognitive model for autism spectrum disorder (ASD). *Cochrane database of systematic reviews*.
[10.1002/14651858.CD008785.pub2](https://doi.org/10.1002/14651858.CD008785.pub2)

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Interventions based on the Theory of Mind cognitive model for autism spectrum disorder (ASD) (Review)

Fletcher-Watson S, McConnell F, Manola E, McConachie H



**THE COCHRANE
COLLABORATION®**

This is a reprint of a Cochrane review, prepared and maintained by The Cochrane Collaboration and published in *The Cochrane Library* 2014, Issue 3

<http://www.thecochranelibrary.com>

WILEY

TABLE OF CONTENTS

HEADER	1
ABSTRACT	1
PLAIN LANGUAGE SUMMARY	2
SUMMARY OF FINDINGS FOR THE MAIN COMPARISON	4
BACKGROUND	8
OBJECTIVES	9
METHODS	9
RESULTS	14
Figure 1.	15
Figure 2.	18
Figure 3.	19
Figure 4.	22
Figure 5.	23
Figure 6.	24
DISCUSSION	25
AUTHORS' CONCLUSIONS	27
ACKNOWLEDGEMENTS	27
REFERENCES	28
CHARACTERISTICS OF STUDIES	35
DATA AND ANALYSES	66
Analysis 1.1. Comparison 1 Treatment effects in meta-analysis, Outcome 1 Joint engagement in mother-child interaction.	66
Analysis 1.2. Comparison 1 Treatment effects in meta-analysis, Outcome 2 Emotion recognition from face photographs, TAU control.	67
Analysis 1.3. Comparison 1 Treatment effects in meta-analysis, Outcome 3 Joint attention initiations in standardised assessment.	68
ADDITIONAL TABLES	68
APPENDICES	73
CONTRIBUTIONS OF AUTHORS	78
DECLARATIONS OF INTEREST	79
SOURCES OF SUPPORT	79
DIFFERENCES BETWEEN PROTOCOL AND REVIEW	80

[Intervention Review]

Interventions based on the Theory of Mind cognitive model for autism spectrum disorder (ASD)

Sue Fletcher-Watson¹, Fiona McConnell¹, Eirini Manola², Helen McConachie³

¹Moray House School of Education, University of Edinburgh, Edinburgh, UK. ²Puzzle - School for Children with Autism, Athens, Greece. ³Institute of Health and Society, Newcastle University, Newcastle upon Tyne, UK

Contact address: Sue Fletcher-Watson, Moray House School of Education, University of Edinburgh, St John's Land, Holyrood Road, Edinburgh, EH8 8AQ, UK. sfwatson@staffmail.ed.ac.uk. sfwatson@staffmail.ed.ac.uk.

Editorial group: Cochrane Developmental, Psychosocial and Learning Problems Group.

Publication status and date: New, published in Issue 3, 2014.

Review content assessed as up-to-date: 7 August 2013.

Citation: Fletcher-Watson S, McConnell F, Manola E, McConachie H. Interventions based on the Theory of Mind cognitive model for autism spectrum disorder (ASD). *Cochrane Database of Systematic Reviews* 2014, Issue 3. Art. No.: CD008785. DOI: 10.1002/14651858.CD008785.pub2.

Copyright © 2014 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ABSTRACT

Background

The 'Theory of Mind' (ToM) model suggests that people with autism spectrum disorder (ASD) have a profound difficulty understanding the minds of other people - their emotions, feelings, beliefs, and thoughts. As an explanation for some of the characteristic social and communication behaviours of people with ASD, this model has had a significant influence on research and practice. It implies that successful interventions to teach ToM could, in turn, have far-reaching effects on behaviours and outcome.

Objectives

To review the efficacy of interventions based on the ToM model for individuals with ASD.

Search methods

In August 2013 we searched CENTRAL, Ovid MEDLINE, Embase, CINAHL, PsycINFO, ERIC, Social Services Abstracts, Autism-Data, and two trials registers. We also searched the reference lists of relevant papers, contacted authors who work in this field, and handsearched a number of journals.

Selection criteria

Review studies were selected on the basis that they reported on an applicable intervention (linked to ToM in one of four clearly-defined ways), presented new randomised controlled trial data, and participants had a confirmed diagnosis of an autism spectrum disorder. Studies were selected by two review authors independently and a third author arbitrated when necessary.

Data collection and analysis

Risk of bias was evaluated and data were extracted by two review authors independently; a third author arbitrated when necessary. Most studies were not eligible for meta-analysis, the principal reason being mis-matching methodologies and outcome measures. Three small meta-analyses were carried out.

Main results

Twenty-two randomised trials were included in the review (N = 695). Studies were highly variable in their country of origin, sample size, participant age, intervention delivery type, and outcome measures. Risk of bias was variable across categories. There were very few studies for which there was adequate blinding of participants and personnel, and some were also judged at high risk of bias in blinding of outcome assessors. There was also evidence of some bias in sequence generation and allocation concealment. Not all studies reported data that fell within the pre-defined primary outcome categories for the review, instead many studies reported measures which were intervention-specific (e.g. emotion recognition). The wide range of measures used within each outcome category and the mixed results from these measures introduced further complexity when interpreting results.

Studies were grouped into four main categories according to intervention target/primary outcome measure. These were: emotion recognition studies, joint attention and social communication studies, imitation studies, and studies teaching ToM itself. Within the first two of these categories, a sub-set of studies were deemed suitable for meta-analysis for a limited number of key outcomes.

There was very low quality evidence of a positive effect on measures of communication based on individual results from three studies. There was low quality evidence from 11 studies reporting mixed results of interventions on measures of social interaction, very low quality evidence from four studies reporting mixed results on measures of general communication, and very low quality evidence from four studies reporting mixed results on measures of ToM ability.

The meta-analysis results we were able to generate showed that interventions targeting emotion recognition across age groups and working with people within the average range of intellectual ability had a positive effect on the target skill, measured by a test using photographs of faces (mean increase of 0.75 points, 95% confidence interval (CI) 0.22 to 1.29 points, $Z = 2.75$, $P < 0.006$, four studies, $N = 105$). Therapist-led joint attention interventions can promote production of more joint attention behaviours within adult-child interaction (mean increase of 0.55 points, 95% CI 0.11 to 0.99 points, $Z = 2.45$, $P \text{ value} = 0.01$, two studies, $N = 88$). Further analysis undermines this conclusion somewhat by demonstrating that there was no clear evidence that intervention can have an effect on joint attention initiations as measured using a standardised assessment tool (mean increase of 0.23 points, 95% CI -0.48 to 0.94 points, $Z = 0.63$, $P \text{ value} = 0.53$, three studies, $N = 92$). No adverse effects were apparent.

Authors' conclusions

While there is some evidence that ToM, or a precursor skill, can be taught to people with ASD, there is little evidence of maintenance of that skill, generalisation to other settings, or developmental effects on related skills. Furthermore, inconsistency in findings and measurement means that evidence has been graded of 'very low' or 'low' quality and we cannot be confident that suggestions of positive effects will be sustained as high-quality evidence accumulates. Further longitudinal designs and larger samples are needed to help elucidate both the efficacy of ToM-linked interventions and the explanatory value of the ToM model itself. It is possible that the continuing refinement of the ToM model will lead to better interventions which have a greater impact on development than those investigated to date.

PLAIN LANGUAGE SUMMARY

A review of evidence on the use of interventions for people with autism spectrum disorder, based on the psychological model 'Theory of Mind'

Background

The 'Theory of Mind' model suggests that people with autism spectrum disorder (ASD) have a profound difficulty understanding the minds of other people, their emotions, feelings, beliefs, and thoughts. It has been proposed that this may underlie many of the other difficulties experienced by people with ASD, including social and communication problems, and some challenging behaviours. Therefore, a number of studies have attempted to teach theory of mind and related skills to people with ASD.

Review question

This review aimed to explore whether a) it is possible to teach theory of mind skills to people with autism and b) whether or not this evidence supports the theory of mind model. Having a 'theory of mind' may depend on developing related basic skills, including joint attention (sharing a focus of interest with another person), recognising other people's emotions from faces or stories, and imitating other people. Therefore, we included intervention studies that taught not just theory of mind itself, but also related skills.

Study characteristics

We found 22 research studies involving 695 participants, which reported on the efficacy of interventions related to theory of mind. The evidence is current to 7th August 2013.

Key results and the quality of the evidence

Despite all studies using a high-quality basic methodology (the randomised controlled trial), there was concern over poor study design and reporting in some aspects. While there is some evidence that theory of mind, or related skills, can be taught to people with ASD, there is currently poor quality evidence that these skills can be maintained, generalised to other settings, or that teaching theory of mind has an impact on developmentally-linked abilities. For example, it was rare for a taught skill to generalise to a new context, such as sharing attention with a new adult who was not the therapist during the intervention. New skills were not necessarily maintained over time. This evidence could imply that the theory of mind model has little relevance for educational and clinical practice in ASD. Further research using longitudinal methods, better outcome measures, and higher standards of reporting is needed to throw light on the issues. This is particularly important as the specific details of the theory of mind model continue to evolve.

SUMMARY OF FINDINGS FOR THE MAIN COMPARISON *[Explanation]*

Theory of Mind based interventions compared with wait-list or treatment-as-usual control for autism spectrum disorder.					
<p>Patient or population: People with autism spectrum disorder Settings: Schools, home and clinical settings Intervention: Based on the Theory of Mind theoretical model of autism Comparison: Most studies incorporate an 'empty' control such as treatment-as-usual or wait-list</p>					
Outcomes	Illustrative comparative risks* (95% CI)		No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk			
	[Control]	[Intervention]			
Symptom Level: Communication Various measures, including: Autism Diagnostic Observation Schedule (ADOS) Conversation Skills Social Communication Questionnaire (SCQ) (level of eye-contact)	See 'Corresponding Risk'	Wong 2010 and Young 2012 report positive effects of intervention on symptom level in the communication domain, while Hadwin 1996 found no effect on conversational skills (this specific outcome is reported in Hadwin 1997)	ADOS: n = 17 (Wong 2010) Conversation: n = 30 (Hadwin 1996) SCQ: n = 25 (Young 2012)	⊕○○○ very low [1]	Three included studies report outcomes in this area of clinical relevance. Each one uses a different assessment to capture change in this domain. One study uses an unstandardised measure, though it is designed to capture change over time (Hadwin 1996). The other two studies use standardised measures of communication skills but neither of these were designed to capture change over time nor to be used as intervention outcome measures
Symptom Level: Social Interaction Various measures, including: Autism Diagnostic Observation Schedule (ADOS)	See 'Corresponding Risk'	Fewer than half of the relevant included studies report positive effects of intervention on symptom level in the social interaction domain	ADOS: n = 17 (Wong 2010) CSBS: n = 48 (Landa 2011) ESCS: n = 200	⊕⊕○○ low [2]	Here we include both standardised assessments and direct observations of social behaviours Eleven included studies report

Communication and Symbolic Behaviour Scale (CSBS)
 Early Social Communication Scales (ESCS)
 PDD-BI social approach subscale
 Precursors of Joint Attention Measure
 Social (PJAM) Communication Questionnaire (SCQ)
 Social Emotional Scale (SES) (Bayley-III)
 Social Skills Rating System (SSRS)
 Vineland Adaptive Behaviour Scales (socialisation) (VABS)
and
 Other social interaction (SI) observations

(Ingersoll 2012; Kasari 2006; Hopkins 2011; Landa 2011; Wong 2010).

In addition some studies report mixed findings across methods. For example, Goods 2013 and Kaale 2012 report some positive effects measured in observations, but null findings from the ESCS. Conversely Kim 2009 (outcomes reported in Kim 2008) and Wong 2013 find significant effects measured by the ESCS but not all other measures. In the case of Wong 2013 this is further complicated by a mixed output from the ESCS where a significant effect is found for one scored item but not another. Similar findings are reported by Schertz 2013 and Kasari 2010 who find positive effects of intervention on some observed behaviours but not others

Both studies which report the impact of an emotion recognition intervention on generalised social skills do not find significant effects on their chosen outcomes (Williams 2012; Young 2012)

(Goods 2013; Ingersoll 2012; Kaale 2012; Kasari 2006; Kim 2009; Wong 2013)

PDD-BI: n = 10
(Kim 2009)

PJAM: n = 23
(Schertz 2013)

SCQ: n = 25
(Young 2012)

SES Bayley: n = 27
(Ingersoll 2012)

SSRS: n = 49
(Hopkins 2011)

VABS Socialisation: n = 55
(Williams 2012)

Other SI: n = 175
(Kasari 2006, Kim 2009; Kaale 2012; Kasari 2010; Goods 2013)

outcomes in this area of clinical relevance. There is wide variety in the choice of assessments to capture change in this domain, though most are based on standardised assessments and are often designed to capture change over time

<p>General Communication Ability (e.g. vocabulary) Mullen Scales of Early Learning (MSEL) Reynell Developmental Language Scales</p>	<p>See 'Corresponding Risk'</p>	<p>Schertz 2013 reports significant intervention effects on receptive language and a non-significant but moderate sized effect ($d = 0.78$) for expressive language scores. At a one-year follow-up Kasari 2006 likewise report intervention effects on expressive language, which were significantly greater for the joint attention intervention compared with both control group and symbolic play interventions. However these effects on expressive vocabulary were not sustained four years later (Kasari 2012b). In addition, a methodologically strong study (Landa 2011) reports no effects on expressive language.</p>	<p>MSEL expressive: $n = 71$ (Landa 2011, Schertz 2013) MSEL receptive: $n = 23$ (Schertz 2013) Reynell: $n = 58$ (Kasari 2006)</p>	<p>⊕○○○ very low [3]</p>	<p>Though this has commonly been used as an outcome measure in generalised social skills interventions for children with ASD, only three of the studies included in this review report a general communication ability outcome measure</p>
<p>Theory of Mind ability Various measures, including: False-belief tasks Happé's Strange Stories Faux-Pas Recognition Test NEPSY-II ToM tasks The ToM Test</p>	<p>See 'Corresponding Risk'</p>	<p>Two studies report some positive effects of intervention on ToM ability (Begeer 2011; Fisher 2005) one reports no impact on directly-assessed ToM ability (Solomon 2004) and one reports a reduction at follow-up in ToM ability for the intervention group specifically (Williams 2012).</p>	<p>False belief: $n = 27$ (Fisher 2005) Happé SS, & Faux-Pas RT: $n = 18$ (Solomon 2004) NEPSY: $n = 55$ (Williams 2012) ToM Test: $n = 36$ (Begeer 2011)</p>	<p>⊕○○○ very low</p>	<p>Four included studies report outcomes in this area of principally theoretical relevance. There is wide variety in the choice of assessments to capture change in this domain, though most are based on standardised assessments and are closely linked to the intervention target skill</p>

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **RR:** Risk Ratio

GRADE Working Group grades of evidence

High quality: Further research is very unlikely to change our confidence in the estimate of effect.

Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.

Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.

Very low quality: We are very uncertain about the estimate.

1. Risk of bias (-1); inconsistency (-2): Since the studies included here are of variable methodological quality and report mixed findings this evidence is considered to be of Very Low quality.
2. Risk of bias (-1); inconsistency (-1): The studies included here are of variable methodological quality and report mixed findings from a wide variety of measures. There is a collection of studies reporting on the ESCS (some of which are summarised in [Analysis 1.1](#)), but within this group findings are once again mixed. Indeed, even within a single study and measure there may be inconsistency in evidence for intervention efficacy. It is therefore impossible to be confident about the impact of Theory of Mind interventions on social interaction domain symptom level and the evidence quality is rated as Low.
3. Risk of bias (-1); inconsistency (-1); low sample size (-1): These mixed outcomes from only a handful of studies must be judged of Very Low Quality until they can be resolved by additional high-quality evidence.

It is challenging to divide communication and social interaction for measures which tap into both of these qualities. However for the purposes of this table, we have identified measures which are based on observation of an interpersonal interaction as falling into the Social Interaction Domain.

A number of included studies report on measures of emotion recognition and imitation skill. While these are suitable outcomes for the respective interventions, and highly associated with ASD profiles, these cannot be categorised into the domains for this Summary of Findings table, and therefore are not addressed here.

BACKGROUND

Description of the condition

Autism spectrum disorder (ASD) is an umbrella term used to describe all people diagnosed as showing symptoms within two core criteria: communication and social deficits, and fixed or repetitive behaviours (APA 2013). The ASD label replaces former subtypes, including autism, pervasive developmental disorder - not otherwise specified (PDD-NOS), and Asperger's syndrome (AS) (APA 1994). Likewise, the single communication and social interaction cluster is derived from what was originally two separate domains of impairment in communication and social interaction (APA 1994). These difficulties often make it very hard for people with ASD to be successful members of society and can present very serious challenges to parents, teachers, and other professionals.

Prevalence estimates of ASD diagnosis in children have been rising significantly in recent years with an authoritative systematic review estimating global prevalence of pervasive developmental disorders at 62 per 10,000 and autistic disorder at 17 per 10,000 (Elsabbagh 2012). Figures may be higher in more developed countries (e.g. Baird 2006). This represents a more-than-threefold increase on previously published figures, which estimated autism prevalence at about 5 per 10,000 (Fombonne 2001). While there are methodological differences between prevalence studies, the rising prevalence of ASD has been well-documented across Western countries, including Europe, Australia, and the USA (e.g. Yeargin-Allsopp 2003; Williams 2006; Atladottir 2007; Kogan 2009; Nassar 2009). There has been significant debate about the cause of the recent rise in prevalence of ASD, but the influence of increased awareness of the disorder among health professionals and the community at large, and the role of diagnostic substitution, should not be underestimated (Croen 2002; Atladottir 2007). There are other candidate explanations, including the possibility of environmental causes of the rising prevalence estimates, though, as yet, there is no good empirical evidence for these (Rutter 2005). Baird et al (Baird 2006) conclude that "Whether the increase is due to better ascertainment, broadening diagnostic criteria, or increased incidence is unclear" (p. 210).

Within the disorder there is a male to female ratio of 4:1 or 5:1 (Baird 2006; Kogan 2009), as noted in the set of case studies, which defined the condition for the first time (Kanner 1943). ASDs have this feature in common with most other neurodevelopmental disorders (such as attention deficit hyperactivity disorder (ADHD), dyslexia, dyspraxia), though to a greater extent. As yet, there is no empirical evidence for systematic differences between male and female individuals with ASD (Hartley 2009).

Theory of Mind

The term 'Theory of Mind' (ToM) describes the ability to understand another's thoughts, beliefs, and other internal states and

was originally applied to the study of non-human primate cognition (Premack 1978). The term has since been developed in a number of different directions (e.g. Carruthers 1996), including in research into ASD. The first application of the term in ASD research was in an experiment which used false-belief paradigms to explore ToM in children with autism (Baron-Cohen 1985). In this study, children were presented with a scenario in which a doll, Sally, 'believed' her marble was in the basket where she left it. However, the child and experimenter knew that while Sally was elsewhere, another doll had moved the marble into a box. The key question was "Where will Sally look for her marble?" Typically-developing children from the age of four years, sometimes earlier, can correctly ascertain that Sally will look in the basket; she holds a false belief about the location of the marble (Wellman 2001). Children with ASD are much less likely to give a correct answer to this question at age four years. They normally claim that Sally will look in the box, in accordance with reality, but incompatible with Sally's knowledge of the situation.

Research into ToM in children and adults with ASD has been prolific over the last 25 years (e.g. Baron-Cohen 2000). While the details are subject to debate, it is widely accepted that people with ASD do not possess a fully-functioning theory of mind; even high-functioning adults with ASD may struggle with complex ToM tasks (Ponnet 2004). ToM has been placed in a developmental context, consisting of a range of precursor skills, including following eye-gaze, establishing joint attention, imitation, pretend play, and emotion recognition (Melzoff 1993; Baron-Cohen 1995; Charman 2000; Wellman 2000; Ruffman 2001). ToM then also links to subsequent social and communication skills, including the development of language (Tager-Flusberg 2000; Garfield 2001). As a result, many believe that failures of ToM are central to explaining the difficulties experienced by people with ASD (though not a sufficient explanation). Therefore, ToM and its precursor skills are targets for interventions.

Description of the intervention

A 'Theory of Mind intervention' is a treatment or therapy, which is explicitly or implicitly based on the Theory of Mind (ToM) cognitive model of ASD. ToM interventions target those skills which are either potential components or precursors of ToM (Swettenham 2000). One example of an intervention targeting such skills is using 'thought-bubbles' to teach children with ASD to understand others' thoughts and beliefs by illustrating these in bubbles (as in a cartoon) (Parsons 1999). Specific precursor skills can also be taught such as helping a child to make eye-contact to accompany pointing to an object of interest (joint attention). More detail on which interventions are eligible for inclusion in this review is given in the Methods section, but we will only consider interventions that explicitly target ToM skills.

ToM interventions can be contrasted with other types of treatment for ASD. Many intervention models focus on behaviour

management and personal skills training, using a basic conditioning model for learning (repetition; rewarding desirable behaviour; 'punishing' or ignoring behaviour that the therapist finds undesirable such as tantrums). In addition, most management strategies for ASD occur within a fairly structured timetable as people with ASD tend to feel more comfortable following familiar routines in a consistent environment, and respond very poorly to change.

How the intervention might work

In a chapter reviewing evidence for the possibility of teaching ToM to individuals with autism, Swettenham states (p. 442) that "a successful method for teaching theory of mind may alleviate the impairments in social interaction that are so debilitating in autism" (Swettenham 2000).

The ToM model of autism suggests that the social and communication difficulties that are characteristic of the syndrome stem from a failure to develop an intact ToM. Certainly there is evidence that ToM is correlated with real-life social skills (Frith 1994) and symptomatology (Joseph 2004). Certain ToM precursor skills also have a direct relationship with symptoms (Mundy 1994). Therefore, training in ToM, or in the precursor or component skills of ToM, should alleviate the social and communication difficulties experienced by individuals with the disorder. For example, a targeted joint attention intervention for autism produced improvements in children's responsiveness to joint attention opportunities and also improved sharing and language (Kasari 2006; Kasari 2008), indicating that ToM interventions may have consequences for wider developmental abilities.

It is possible that interventions targeting different ToM skills will produce varied types of change in participants, and the extent of change may vary. The method of delivery of the intervention may also produce different outcomes. For example, one might expect an intervention delivered by a trained therapist to have greater impact than one delivered by parents. An intervention taught in school may have a different impact to one delivered in the home. The duration of the intervention may also be significant. Deficits in ToM and related skills vary with age (Happé 1995), IQ (Ozonoff 1991a; Happé 1994; Bowler 1997), specific diagnosis (Ozonoff 1991b; Bowler 1992) and verbal ability (Happé 1995; Garfield 2001). As a result, the specific skill being targeted, the method of intervention delivery, its duration and individual differences between participants in ToM intervention studies will be important factors for consideration and for statistical analysis in this review.

Why it is important to do this review

To date, there is no comprehensive review of ToM interventions for autism, despite the fact that the first study attempting to teach ToM to individuals with autism was published in 1995 (Ozonoff 1995). This review will be of relevance to both the clinical and

academic research communities, since ToM interventions not only have the potential to benefit people with ASD, but also provide a unique and rigorous way to test the theoretical model on which they are based.

OBJECTIVES

To assess the effect of interventions, based on the Theory of Mind (ToM) model, for autism spectrum disorders (ASD), on symptoms in the core diagnostic domains of social and communication impairments in autism, and on language and ToM skills. In addition, in so-doing, to test the applied value of the ToM model of autism.

METHODS

Criteria for considering studies for this review

Types of studies

Randomised and quasi-randomised trials (defined as trials in which allocation was made by, for example, alternate allocation or allocation by date of birth).

Types of participants

Participants of any age with a diagnosis of an ASD, including autism, atypical autism, Asperger's syndrome, and PDD-NOS, according to either ICD-10 (International Classification of Diseases), DSM-IV or DSM-V (Diagnostic Statistical Manual of Mental Disorders) criteria. All diagnostic categories could be included since the validity of differentiating between categories on the spectrum is not well established (Klin 2005). Furthermore, the ToM cognitive model does not distinguish, on a qualitative basis, between different forms of ASD. Participants must have received a 'best estimate' clinical diagnosis, confirmed by the study authors. That is, at a minimum, diagnosis by a multidisciplinary clinical team using standard procedures with reference to the international classification systems. Use of a particular diagnostic tool, such as the Autism Diagnostic Observation Schedule (ADOS) (Lord 2000) or the Autism Diagnostic Interview (ADI-R) (Lord 1994), was desirable but not required. Co-morbid cases were also eligible for inclusion since these individuals are just as needful of intervention for their specifically autistic difficulties.

Types of interventions

Interventions eligible for inclusion in this review:

1. explicitly state that they are designed to teach ToM; or
2. explicitly state that they are designed to teach precursor skills of ToM; or
3. explicitly state that they are based on or inspired by ToM models of autism; or
4. explicitly state that they aim to test the ToM model of autism.

We reiterate that ToM (theory of mind) describes the ability to understand another's thoughts, beliefs, and other internal states and is encapsulated in a test of false belief. Prior to the development of false-belief understanding (at about four years old in typical development), associated precursor skills are in evidence such as joint attention, imitation, and emotion recognition. Relevant interventions include those which explicitly teach children to understand others' mental states (e.g. using visual representations of mental states [McGregor 1998](#)) and those which use naturalistic teaching to develop imitation skills ([Heimann 2006](#)).

The following kinds of interventions are not included in this review:

1. interventions which do not meet the criteria given above;
2. medical interventions (e.g. risperidone for aggression in ASD);
3. dietary interventions (e.g. gluten-free and casein-free diets);
4. interventions which target a particular behaviour rather than a cognitive skill (e.g. over-sensitivity to light modified using colour spectacles; sleep difficulties modified using applied behavioural analysis);
5. language-focused interventions (e.g. to make requests using the Picture Exchange Communication System or spoken single words);
6. interventions which have a broad-base both in terms of methods (e.g. combining computerised learning with parent training and social skills groups) and targets (i.e. addressing a range of social communication skills, some which are ToM-linked but also more general skills such as turn-taking, friendship skills, and conversation).

ToM interventions are compared with the following conditions, where these are used:

1. treatment-as-usual/wait-list control;
2. 'placebo' interventions, for example a 'contact control' such as watching Thomas the Tank Engine DVDs (e.g. [Young 2012](#));
3. intervention with no therapeutic content, (e.g. group leisure activities ([Baghdadli 2013](#))).

All 'doses' (that is the number and length of treatment sessions per week), durations, and methods were eligible for inclusion.

Types of outcome measures

Outcome measures do not form part of the criteria for inclusion of studies in this review.

Primary outcomes

Primary outcomes at a participant symptom level, measured using standardised diagnostic assessments or clinical report. Outcomes will be in each of two symptom domains that have until recently been used in clinical diagnosis and are followed by most diagnostic tests for autism. These are as follows, with examples of outcomes in each category as measured by the ADOS ([Lord 2000](#)) or ADI ([Lord 1994](#)).

1. Communication: overall level of non-echoed language; stereotyped or idiosyncratic use of words or phrases; pointing; gestures; conversation.

2. Social function: unusual eye-contact; facial expressions directed to others; spontaneous initiation of joint attention; shared enjoyment in interaction; quality of rapport.

The third diagnostic domain of Restricted and Repetitive Behaviours (imaginative play or creativity; unusual sensory interests; unusually repetitive interests or stereotyped behaviours; compulsions or rituals) is not included as an expected primary outcome.

Secondary outcomes

In addition, the following secondary outcomes will be included. PARTICIPANT, direct measurement

- Intervention-specific: change in targeted cognitive skill such as false-belief understanding
- Change in participant behaviour or quality of interpersonal interaction, or both, measured by direct observation.

PARENT, teacher (or other individual in caring or educational relationship to the participant) report

- Change in participant behaviour and skills or deficits such as: adaptive skills; school success; challenging behaviours; social participation measured by parent, teacher or other report
- Acceptability of intervention (time, cost)

OTHER

- Intervention process measures e.g. rate of drop-out
- Economic data e.g. financial cost of intervention; time commitment required

Main outcomes for 'Summary of findings' table

The following outcomes measures are specified for a 'Summary of findings' table:

1. symptom level, communication domain;
2. symptom level, social interaction domain;
3. general communication ability (e.g. vocabulary);
4. 'Theory of Mind' ability (e.g. false-belief test score).

Where data were available, we planned to organise outcomes into three time points: immediately post-treatment; medium-term outcome (up to six months post-treatment); and long term (more than six months post-treatment).

The [Summary of findings for the main comparison](#) reports on these outcomes and also includes an estimate of the quality of evidence in each category.

Search methods for identification of studies

The complex nature of ToM interventions makes them difficult to capture adequately using search terms. Therefore, to avoid missing relevant studies, we used a highly sensitive search strategy with just two concepts: the condition (ASD) and a search filter to find RCTs. The core search strategy was developed in Ovid MEDLINE and uses the Cochrane highly sensitive search strategy for identifying randomised trials (Lefebvre 2008). The MEDLINE strategy was adapted for other databases using appropriate syntax and controlled vocabulary. The initial searches were run in July 2010 without any date or language restrictions. We last updated the searches on 6 August 2013, apart from ASSIA which was no longer available to us.

Electronic searches

We searched the following databases in August 2013.

- Cochrane Central Register of Controlled Trials (CENTRAL) 2013, Issue 7, part of The Cochrane Library.
- Ovid MEDLINE(R) 1946 to July Week 4 2013.
- EMBASE 1980 to 2013 Week 31.
- CINAHLPlus 1937 to current.
- PsycINFO 1806 to July Week 5 2013.
- ERIC 1966 to current.
- Applied Social Sciences Index and Abstracts: ASSIA (CSA) 1987 to current.
- Social Services Abstracts 1979 to current.
- metaRegister of Controlled Trials (controlled-trials.com/mrct/).
- ICTRP (apps.who.int/trialsearch/).
- UKCRN - UK Clinical Trials Network (public.ukcrn.org.uk/search/).
- ClinicalTrials.gov (clinicaltrials.gov/).
- Autism Data (autism.org.uk/autismdata/).

The search strategies for each source are in [Appendix 1](#).

Searching other resources

In addition to searches of electronic databases, we contacted key authors in the field directly and asked them to provide any relevant published, unpublished or in-progress data, including post-graduate dissertations. We also searched the bibliographies of key articles for citations of papers not found electronically. Searches were made for in-progress, or unpublished clinical trials. Finally, we searched the online databases of journals that regularly publish work on this topic. These journals were the *Journal of Autism and Developmental Disorders*, *Journal of Child Psychology and Psychiatry*, and *Autism: International Journal of Research and Practice*. We also searched the proceedings of the *International Meeting for Autism Research*.

Data collection and analysis

Selection of studies

All citations sourced from the search strategy were transferred to EndNote, a reference management programme. Initial screening of titles and abstracts by an experienced research assistant (EM or FMcC) eliminated all those citations obviously irrelevant to the topic, for example, prevalence studies, studies unrelated to ASD, and single case studies. Thereafter, two review authors (SFW and either EM or FMcC) assessed and selected studies for inclusion from the group of superficially relevant studies. In the event of a disagreement, resolution was reached in discussion with a third author (HM), if necessary following inspection of the full paper.

Data extraction and management

Two review authors (SFW and either EM or FMcC) independently extracted data from selected trials using a specially designed data extraction form. Extracted data included methods (dose and frequency of intervention); diagnostic description of participants, and type of intervention, including target, intensity, duration, and method of application (parent-mediated, therapist, school-based etc.). Disagreements were resolved in consultation with a third author (HM).

Assessment of risk of bias in included studies

Two review authors (SFW and either EM or FMcC) independently assessed the risk of bias of included studies in the following domains: sequence generation; allocation concealment; blinding of participants and personnel; blinding of outcome assessment; incomplete outcome data; selective outcome reporting; other sources of bias. We used The Cochrane Collaboration tool for assessing risk of bias in these areas. The process involved recording the appropriate information for each study (e.g. describing the method used to conceal allocation in detail) and evaluating whether there was risk of bias in that area (e.g. was allocation adequately concealed?). Any disagreement was resolved by referral to a third author (HM).

We contacted authors to supply missing information from 16 included studies (Bolte 2002; Solomon 2004; Fisher 2005; Golan 2006; Kasari 2006; Kim 2009; Golan 2010; Kasari 2010; Ryan 2010; Wong 2010; Hopkins 2011; Ingersoll 2012; Young 2012; Baghdadli 2013; Schertz 2013; Wong 2013) and information was received from the majority of authors with the exception of (Solomon 2004; Fisher 2005; Golan 2006; Golan 2010; Wong 2010; Baghdadli 2013).

Studies were allocated to categories according to our evaluation of each area or potential risk of bias as follows:

Random sequence generation

Low risk of bias: adequate sequence generation as indicated by reference to, e.g., random number table, coin tossing, shuffled cards or envelopes, throwing dice, drawing lots.

Unclear (or moderate) risk of bias: indicates uncertainty about whether the sequence was randomly generated.

High risk of bias: a non-random component is described such as sequence generation by odd or even date of birth, by geographical location or by date of entry to the study.

Allocation concealment

Low risk of bias: participants and investigators enrolling participants unable to foresee assignment as indicated by reference to, e.g., central allocation, opaque envelope procedure, allocation by an independent partner outside the research team.

Unclear (or moderate) risk of bias: indicates uncertainty about whether the allocation was concealed.

High risk of bias: participants and investigators enrolling participants may have been able to foresee assignment as indicated by reference to an open random allocation schedule (e.g. random numbers list), unsealed or non-opaque envelopes, alternate allocation, allocation by non-random criteria such as date of birth.

Blinding of participants and personnel

Low risk of bias: participants and personnel blinded to study hypotheses and treatment condition, or incomplete blinding but authors judge that outcome is unlikely to be influenced by lack of blinding. Examples might be when participants are very young and/or low-functioning people with autism and are unlikely to be aware of intervention targets, and where outcome is assessed using a measure resilient to performance bias such as computerised assessment. We note that in behavioural studies, such as those included in this review, it is rarely possible to blind participants and/or personnel.

Unclear (or moderate) risk of bias: indicates uncertainty about whether blinding was consistent, perhaps due to insufficient information being available, or partial blinding (e.g. of participants but not personnel).

High risk of bias: participants and personnel not blind to study hypotheses or treatment condition, and outcome likely to be influenced by this lack of blinding.

Blinding of outcome assessment

Low risk of bias: outcome examiners and scorers blinded to participant group membership, or blinding of some outcome assessors with good evidence of agreement between blinded and unblinded raters on outcome measures, or outcome assessors not blind but outcome measurement unlikely to be influenced by this lack of blinding.

Unclear (or moderate) risk of bias: indicates uncertainty about whether blinding was consistent, perhaps due to insufficient information being available.

High risk of bias: outcome examiners and scorers not blind to participant group membership, and outcome likely to be influenced by this lack of blinding.

Incomplete outcome data

Low risk: no missing data, or reasons for missing outcome data unlikely to be related to true outcome, or missing data balanced across groups with similar reasons in each case.

Unclear risk of bias: insufficient reporting of attrition or exclusions to permit accurate judgement.

High risk of bias: reasons for missing data likely to be related to true outcome, with imbalance in numbers between groups or different reasons between groups.

Selective reporting

Low risk: study protocol available and all pre-specified outcomes are reported in the pre-specified way, or clear from the published reports that all expected outcomes are included.

Unclear (moderate) risk of bias: insufficient information to permit accurate judgement.

High risk of bias: not all of the study's primary outcomes have been reported, or outcomes which were not pre-specified are reported, or one or more primary outcomes have been reported for only a sub-set of the sample, or one or more outcomes are reported incompletely so that they cannot be entered into a meta-analysis.

Other bias

Low risk of bias: the study appears to be free of other sources of bias.

Unclear (or moderate) risk of bias: there may be an additional risk of bias but there is insufficient information to fully assess this risk, or it is unclear whether the risk would introduce bias in study results.

High risk of bias: the study has one important additional risk of bias such as a source of bias related to the study design, or claims of fraudulence.

In each case, only studies where the assessment of overall risk falls into categories 'Low' or 'Unclear/Moderate' have been included in subsequent analyses.

Measures of treatment effect

Binary and categorical data

No studies reported binary outcome data in the current review version. Should they be included in future updates, methods for

analysing them appear in the published protocol for the review (Fletcher-Watson 2010).

Continuous data

Where standardised assessment tools generated a continuous score as the outcome measure, and means and standard deviations were reported or provided by the authors, comparisons were made between the means of these scores. When selecting studies for possible meta-analysis the following criteria were of principal importance.

- Similarity of trial design - especially whether the 'control' condition had therapeutic content or not
- Similarity of intervention target
- Similarity of outcome measure - the quality being measured, the unit of measurement, and the method of measurement (e.g. parent-report, video coding, standardised assessment)

Where measures were on different scales but those scales were clinically homogeneous, meta-analyses used standardised mean difference with Hedges' *g* correction for small sample sizes.

Unit of analysis issues

No cluster-randomised trials were included in the current review version. Methods for analysis are recorded in additional Table 1.

Dealing with missing data

Missing data were assessed for each individual study according to the reports provided by authors. For included studies reporting drop-out, we reported the number of participants included in the final analysis as a proportion of those participants who began the intervention (see [Characteristics of included studies](#)). Reasons for missing data are also reported (that is, whether data are missing at random or not). In all cases, we concluded that data were missing at random, and the remaining data were analysed and the missing data ignored.

Where summary data are missing, trial authors were contacted. If no reply was forthcoming or the required summaries were not made available, the study was included in the review and we assessed and discussed the extent to which its absence from meta-analysis affects the review results (e.g. Bolte 2002).

No studies reported the loss of significant quantities of data, without sufficient explanation, and there was no evidence of non-random missing data. Therefore, the review authors agree that the conclusions of individual studies are not compromised by missing data. The extent to which the results of the review may be altered by the missing data is assessed and discussed ([Quality of the evidence](#)).

Additional procedures for dealing with non-random missing data in future appear in the published review protocol (Fletcher-Watson 2010) and Table 1.

Assessment of heterogeneity

Consistency of results was assessed visually and by a Chi^2 test. Where meta-analysis included only a small number of studies, or where studies had small sample sizes, a *P* value of 0.10 was applied for statistical significance. In addition, since Chi^2 can have low power when only a few studies or studies of a small sample size are available, we used the I^2 statistic to measure the amount of observed variability in effect sizes that can be attributed to true heterogeneity (Higgins 2008).

Assessment of reporting biases

Where sufficient studies were found, funnel plots were inspected to investigate any relationship between effect size and sample size. Such a relationship could be due to publication or related biases, or due to systematic differences between small and large studies.

Data synthesis

Data synthesis was performed using RevMan 5.2. Binary data were not reported in any of the included studies but could be assessed in future review versions. Where two or more studies suitable for inclusion were found, and the studies were considered to be homogenous, a meta-analysis was performed on the results. Homogeneity decisions were based on examination of a series of factors identified in the review protocol including the following.

- Similarity in intervention delivery type (e.g. therapist-led, parent training)
- Similarity in intervention target skill (e.g. emotion recognition, imitation, joint attention)
- Similarity in participant populations (e.g. intellectual level in the normal or low range, specific autism diagnostic category, age)

In addition, the following two further factors were developed post hoc in response to the wide variability in study design and outcome measure found.

- Similarity in primary outcome measurement
- Similarity in comparison group status (e.g. did the study compare two different interventions or compare an intervention with a wait-list or treatment-as-usual control)

It is essential to distinguish between measures of primary outcome when assessing intervention efficacy for two main reasons. The first is that there is significant evidence that people with ASD do not generalise skills across contexts. For example, Golan 2010 found differences in outcome measures by close and distant generalisation tasks even though these were all measures of emotion and mental state recognition. Therefore, studies measuring outcome using tasks which differ in complexity and in connection to the teaching context should not be compared directly. The second reason is that the method of measurement can produce widely varying distributions, which are not amendable to combination. For example, Kasari 2010 measured percentage of total time that a mother-

child dyad were jointly engaged, while [Goods 2013](#) reported the number of instances that specific types of joint engagement were observed. It is not possible to combine these two variables, which have also been collected in different settings (laboratory artificial mother-child play versus naturalistic classroom observation) and over different periods of time.

Comparison group status is another key consideration when combining studies. A study that shows an intervention effect compared with a 'placebo' group or compared with another intervention may have a smaller effect size than one comparing intervention and wait-list control. However, the former study has the more powerful design and so this smaller effect should be more influential on conclusions.

A random-effects model analysis was used since we do not assume that each study is estimating exactly the same quantity.

Subgroup analysis and investigation of heterogeneity

Subgroup analyses were not possible in this version of the review. Dimensions for possible future subgroup analyses are included in additional [Table 1](#).

Sensitivity analysis

Sensitivity analysis was not possible for this version of the review. Details of planned future sensitivity analyses are included in additional [Table 1](#).

RESULTS

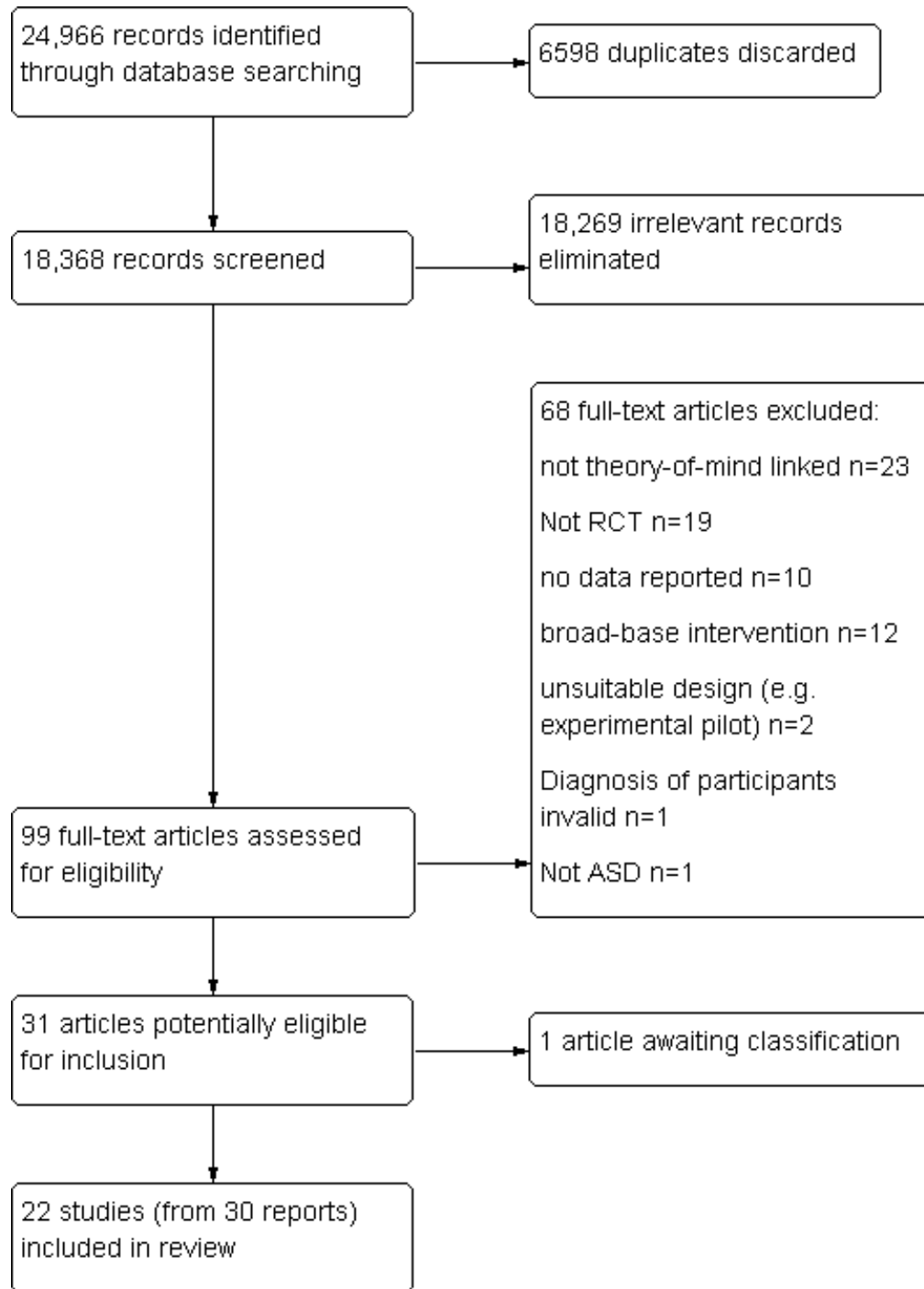
Description of studies

See: [Characteristics of included studies](#), [Characteristics of excluded studies](#).

Results of the search

Searches were carried out in July 2010, and again in July 2012, and August 2013, yielding 18,368 records of potential relevance after de-duplication (July 2010: 11,822 records, July 2012: 4171 records, August 2013: 2375 records). Assessment of titles and abstracts and elimination of duplicates between the searches resulted in a list of 99 records for closer examination ([Figure 1](#)). One of these articles is only available in French and is currently awaiting classification pending translation ([Baghdadli 2010](#)).

Figure 1. Study flow diagram



Included studies

We included 22 studies involving 695 participants in this review; in each case the main study is reported in a published journal article and the dates of publication span from 1996 to 2013. There were 17 studies reported in a single published journal article: Solomon 2004; Fisher 2005; Golan 2006; Golan 2010; Kasari 2010; Landa 2011; Ryan 2010; Wong 2010; Begeer 2011; Hopkins 2011; Kaale 2012; Williams 2012; Young 2012; Baghdadli 2013; Goods 2013; Schertz 2013; Wong 2013. In the case of Golan 2006, two studies are reported of which only Experiment One is an RCT, and therefore only this first data set is included in the review.

In addition, there are five studies for which data have been reported in multiple outputs. These are a therapist-led theory of mind intervention (Hadwin 1996), a computerised emotion recognition intervention study (Bolte 2002), an imitation intervention study (Ingersoll 2012), a music therapy study (Kim 2009), and a joint attention and symbolic play intervention conducted by Kasari and colleagues (Kasari 2006). In the case of the Kasari study, one output is an unpublished PhD thesis (Arora 2008).

All 22 studies described themselves as randomised controlled trials, and they were conducted in a wide variety of locations: Scandinavia (Bolte 2002; Kaale 2012); mainland Europe (Begeer 2011; Baghdadli 2013); the UK and Ireland (Hadwin 1996; Fisher 2005; Golan 2006; Golan 2010; Ryan 2010); the Far East (Kim 2009; Wong 2010); Australia (Williams 2012; Young 2012); and the USA (Solomon 2004; Kasari 2006; Kasari 2010; Landa 2011; Hopkins 2011; Ingersoll 2012; Goods 2013; Schertz 2013; Wong 2013).

Participant baseline characteristics

Participants varied widely in age-range from preschoolers (e.g. Kasari 2006) to adolescents and adults (e.g. Bolte 2002) but a majority focused on either pre-school or primary-school aged children (see [Characteristics of included studies](#)). Almost all studies included both boys and girls, though the proportion of male participants was much higher than females, corresponding to the known greater prevalence of diagnosed ASD in males (Kogan 2009). Four studies reported an all-male sample (Bolte 2002; Solomon 2004; Kim 2009; Baghdadli 2013).

For all studies, a diagnosis of an ASD was a requirement for inclusion. A large proportion confirmed diagnosis using a clinical instrument such as the Autism Diagnostic Observation Schedule (ADOS, Lord 1994) or the Childhood Autism Rating Scale (CARS, Schopler 1986). Two studies accepted prior clinical diagnosis as adequate (Hadwin 1996; Fisher 2005), but these also instituted a checklist confirming that all diagnostic criteria were met. Participants were reported as having a range of ASD diag-

noses, including autism, autism spectrum disorder, pervasive developmental disorder - not otherwise specified (PDD-NOS), high-functioning autism (HFA), and Asperger's syndrome (AS). Studies recruiting participants with HFA and/or AS had participants in the adolescent and adult age-range (e.g. Bolte 2002; Golan 2006) or late childhood (Solomon 2004; Begeer 2011). Studies with young children and preschoolers largely described participants as having 'core' autism, or ASD.

All studies reported some measure of general intellectual ability such as verbal mental age. Almost half of the included studies included a sample in the normal intellectual range (Bolte 2002; Solomon 2004; Golan 2006; Kim 2009; Golan 2010; Ryan 2010; Begeer 2011; Young 2012; Baghdadli 2013) and the rest reported on a sample with intellectual disability. One study split the participant group into those with and without associated intellectual delay (Hopkins 2011).

Sample sizes varied widely from $n = 10$ (Bolte 2002; Kim 2009) to $n = 61$ (Kaale 2012). On the whole, very small proportions of participants failed to complete the interventions. The maximum drop-out rate was 27% from a small sample (Goods 2013), but many studies reported no drop-out at all.

Intervention target types

The reported intervention types can be assigned to the following categories, taken from the review protocol.

1. Interventions that explicitly state that they are designed to teach ToM = (Hadwin 1996; Solomon 2004; Fisher 2005; Begeer 2011; Baghdadli 2013).
2. Interventions that explicitly state that they are designed to teach precursor skills of ToM = (Bolte 2002; Golan 2006; Kasari 2006; Kim 2009; Golan 2010; Kasari 2010; Landa 2011; Ryan 2010; Wong 2010; Hopkins 2011; Ingersoll 2012; Kaale 2012; Williams 2012; Young 2012; Goods 2013; Schertz 2013; Wong 2013).
3. Interventions that explicitly state that they are based on or inspired by ToM models of autism.
4. Interventions that explicitly state that they are designed to test the ToM model of autism.

There were no studies falling into category three or four and the vast majority of studies stated that they were designed to teach precursor skills of ToM. Within this category we could also identify some common intervention targets including the following.

- Emotion recognition (Bolte 2002; Golan 2006; Golan 2010; Ryan 2010; Hopkins 2011; Williams 2012; Young 2012)
- Joint attention and social communication (Kasari 2006; Kim 2009; Kasari 2010; Landa 2011; Wong 2010; Kaale 2012; Goods 2013; Schertz 2013; Wong 2013)
- Imitation skills (Ingersoll 2012)

Delivery methods, durations and dose

Three studies reported on the use of a computer program to deliver the intervention (Bolte 2002; Golan 2006; Hopkins 2011) and all of these studies had emotion recognition as the target skill. Three studies investigated the effect of a set of specially-designed cartoons on emotion recognition (Golan 2010; Williams 2012; Young 2012). Other studies investigated the effects of one-to-one therapist-led interventions (Hadwin 1996; Fisher 2005; Kasari 2006; Landa 2011; Ryan 2010; Wong 2010; Ingersoll 2012; Goods 2013) and two of these used the same manualised treatment program (Kasari 2006; Kaale 2012). Some used a therapist-led approach in a group treatment setting (Solomon 2004; Begeer 2011; Baghdadli 2013) and one was a group music therapy approach (Kim 2009). Non-expert intervention delivery was rare with only four studies reporting a parent-training element (Solomon 2004; Kasari 2010; Begeer 2011; Schertz 2013) and one study reporting on teacher-training for intervention delivery in the classroom (Wong 2013).

Intervention durations varied widely from two or three weeks (Hadwin 1996; Young 2012) to six months (Landa 2011). Dose was more consistent, with most falling within a range of 30 minutes per week (Kim 2009) to 3.5 hours per week (Hadwin 1996; Kasari 2006; Golan 2010), and one outlying intervention which reported therapist contact time of 2.5 hours per day (Landa 2011).

Most studies had wait-list or treatment-as-usual control conditions. Six studies (Kim 2009; Landa 2011; Hopkins 2011; Williams 2012; Young 2012; Baghdadli 2013) included control conditions, which were not expected to have an impact on intervention outcome but were included as a contact control only. These included toy play, non-synchronous one-to-one time, using art software, group leisure activities, and watching a Thomas the Tank Engine DVD.

Outcome measures

On the whole, studies rarely identified a single primary outcome measure. Those that organised outcomes into primary and secondary categories usually had multiple measures in each category. The outcome measures used most commonly included the following.

- Recognition of emotion from a variety of stimuli, including static images of faces, static images of the eyes, film clips, short stories, and cartoons (Bolte 2002; Solomon 2004; Golan 2006; Golan 2010; Ryan 2010; Hopkins 2011; Williams 2012; Young 2012; Baghdadli 2013)
- Joint attention and joint engagement behaviours, often measured using video coding of parent-child or teacher-child interactions (Kasari 2006; Kim 2009; Kasari 2010; Landa 2011; Ingersoll 2012; Kaale 2012; Goods 2013; Schertz 2013; Wong 2013)
- Direct assessment of ToM abilities (Hadwin 1996; Solomon 2004; Fisher 2005; Begeer 2011; Williams 2012)

- Imitation skills (Landa 2011; Ingersoll 2012)
- Diagnostic outcome (Wong 2010, Young 2012)

The studies below included the following additional outcome measures.

- Caregiver measures such as quality of involvement, adherence to treatment, mental health or satisfaction surveys (Solomon 2004; Kasari 2010; Landa 2011, Wong 2010; Baghdadli 2013; Wong 2013)
- General social skills measures, including rating scales and observation (Kim 2009; Begeer 2011; Hopkins 2011; Ingersoll 2012; Williams 2012)
- Symbolic play measures (Hadwin 1996; Wong 2010) or assessments of play variety (Goods 2013; Wong 2013)
- Language (Kasari 2006; Landa 2011) and conversational skills (Hadwin 1996)
- fMRI (functional magnetic resonance imaging - assessment of brain activity in facial recognition areas) (Bolte 2002)
- Adaptive function (Kasari 2006; Schertz 2013) and general intellectual abilities (Landa 2011; Schertz 2013)

Selection for meta-analyses

Using protocol criteria, three groups of studies were identified as eligible for meta-analysis.

1. Emotion recognition studies, with a treatment-as-usual control, and outcome measures using judgements of emotional expressions from static photographs of faces (Analysis 1.2).
2. Joint attention and social communication studies, with a treatment-as-usual control, and outcome measures using coding of parent-child interaction videos (Analysis 1.1).
3. Joint attention studies, with a treatment-as-usual control, and outcome measures of joint attention initiation frequency within a standardised assessment (the Early Social Communication Scales) (Analysis 1.3).

Excluded studies

Examination of the abstracts and, where necessary, full texts of reports resulted in a number of exclusions, listed in [Characteristics of excluded studies](#) for the following reasons.

- Not fitting the ToM-linked criteria for inclusion (23 reports)
- Not presenting any new data (10 reports)
- Not randomised or quasi-randomised controlled trials (18 reports)
- Not reporting on a sample of people with ASD (one report)
- Reporting on a broad-based intervention without a specific ToM-linked focus (12 reports)
- Diagnosis of participants invalid (one report)
- Reporting on an experimental pilot RCT with a very short intervention period (three reports)

Risk of bias in included studies

Further information was requested from the majority of authors as papers were not always complete in their reporting. The summaries of 'Risk of bias' judgements are shown in [Figure 2](#) and [Figure 3](#).

Figure 2. 'Risk of bias' graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.

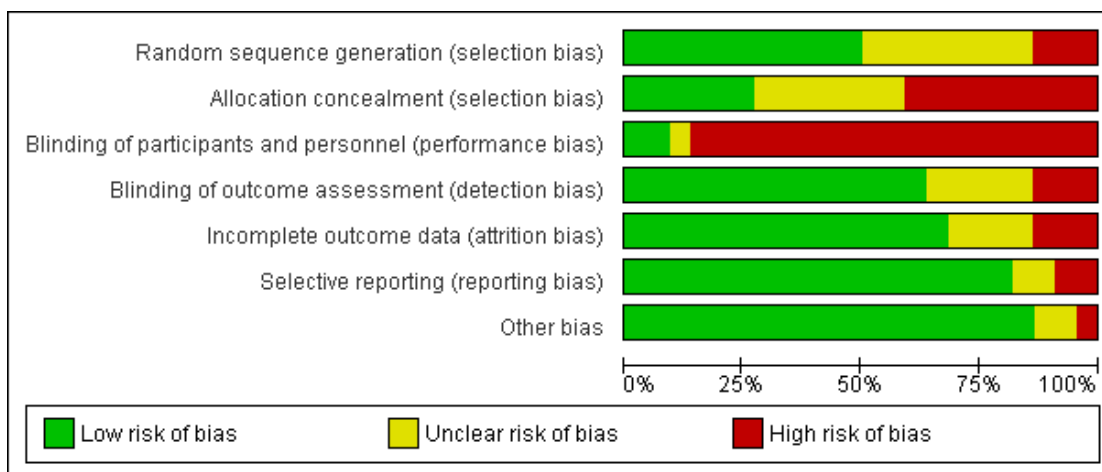


Figure 3. 'Risk of bias' summary: review authors' judgements about each risk of bias item for each included study.

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other bias
Baghdadli 2013	+	?	-	+	+	+	+
Begeer 2011	?	?	-	?	?	+	+
Bolte 2002	+	+	-	?	+	-	+
Fisher 2005	-	-	-	-	?	+	+
Golan 2006	?	-	-	+	-	+	+
Golan 2010	?	?	-	?	+	+	+
Goods 2013	+	-	-	+	-	-	+
Hadwin 1996	?	-	-	-	+	+	+
Hopkins 2011	?	?	?	+	+	+	+
Ingersoll 2012	+	+	-	+	+	+	?
Kaale 2012	+	+	-	+	+	+	+
Kasari 2006	+	-	-	+	?	?	+
Kasari 2010	+	+	-	+	+	+	+
Kim 2009	?	?	-	+	?	+	+
Landa 2011	?	?	+	+	+	+	+
Ryan 2010	+	-	-	?	+	+	+
Schertz 2013	?	-	-	+	+	?	-
Solomon 2004	+	?	-	-	+	+	+
Williams 2012	+	+	-	+	-	+	+
Wong 2010	-	-	-	+	+	+	?
Wong 2013	+	+	-	+	+	+	+
Young 2012	-	-	+	?	+	+	+

Allocation

All studies in this review described themselves as randomised controlled trials and were included on this basis. One was later revealed to have used a non-random allocation procedure (Young 2012). This study states that participants were “randomly allocated to two groups” (Young 2012, p. 986) but email correspondence to clarify the exact allocation procedure revealed that in fact this study used alternate allocation by study enrolment. In other studies, a similar statement is made but rarely is full detail provided.

Thus, only half of the included studies (11 reports) were judged to have ‘low’ risk of bias in terms of the description of the method of randomisation. Only six were judged to have adequately described allocation concealment. Therefore, the majority of included studies have either ‘unclear’ or ‘high’ risk of bias in this category.

In some cases, efforts were made to conceal allocation, for example, using randomisation within blocks to ensure random allocation and smooth delivery of the intervention (Goods 2013). However, the use of blocks of fixed length meant that the final case within each block would be allocated to a known condition.

Blinding

The majority of studies were judged at high risk of bias in this category (19/22 studies, 86%). The three exceptions are Landa 2011 and Hopkins 2011 where partial blinding was achieved, and Young 2012 who created a study design with full blinding as the intervention was delivered not by a therapist but on a DVD.

Blinding of participants and personnel was rarely possible in the studies included in this review, as behavioural interventions were being used and these were often therapist-led. Blinding of outcome assessors is easier to achieve and 14 studies (64%) clearly reported blinding at this stage, though in a further five cases it was unclear whether this was completed adequately.

Though risk of bias must be judged as high when blinding is not achieved, a number of mitigating factors might help to reduce the impact of this risk.

1. When working with very young children or those severely affected by autism and/or intellectual disabilities, it is reasonable to judge that participants are relatively oblivious to the intervention content and certainly to the expected outcomes.
2. Likewise, although participants and parents may be aware of their group they may not be apprised of the hypotheses of the study. For example, Golan 2006 worked with able adults with autism who were asked to “help in the evaluation of a piece of new software” (p. 600) rather than being told the software was designed to help them learn to understand emotions.
3. Many studies used automated outcome measures, especially when using a computerised intervention (e.g. Bolte 2002; Golan 2006), which are more resilient to bias than experimenter-led

methods.

4. Studies using multiple outcome measures often achieved blinding for a sub-set of those outcomes (e.g. Hopkins 2011).

Incomplete outcome data

There was very little evidence of attrition among the studies reported here, and only three of the included studies were judged to be at high risk of bias. The most extreme case of likely bias was (Goods 2013) who reported 73% retention and analysed outcome data for intervention completers only. Ten studies (45%) reported outcome data for all of the original sample and where there was participant drop-out this was usually described with clear reasons to help the reader judge the impact of this drop-out. Where studies were judged at unclear risk of bias, this was due to either a lack of sufficient detail in the published report (e.g. Begeer 2011) or because it was difficult to evaluate the impact of the drop-out on findings (e.g. Kim 2009; Kasari 2006).

Selective reporting

Selective reporting was not evident among the papers included and 18 studies (82%) were judged at low risk of bias in this category. However, it must be noted that the tendency not to identify a primary outcome measure and to use multiple outcomes does hinder conclusions about intervention efficacy.

One study (Bolte 2002) did not report means for a relevant outcome measure - the International Affective Pictures System or IAPS - for which there was no significant group difference. The study authors were contacted to provide mean scores but were unable to provide these due to the time elapsed since the study and records not having been kept. One further study (Ryan 2010) reported data in graphical form only, but the authors kindly provided accurate means and SDs for these data. Conversely, Goods 2013 reported non-significant findings in a table of results but these were not discussed in the text.

Other potential sources of bias

One study reported a significant difference in ratings provided by mothers and ratings provided by independent examiners (Kim 2009) with professional ratings providing a more positive estimate of intervention efficacy. This bias is judged to be of low risk for two reasons. First, the professional ratings used to construct the primary outcome for the intervention were blind to group. Second, the authors provided a reasonable justification for the under-estimation of intervention efficacy by parent ratings, which is that mothers over-estimate pre-intervention abilities of their children, and thus under-estimate efficacy of the intervention.

Another study combined the wait-list control group with the intervention group to provide a larger sample size for analysis of some intervention efficacy measures (Wong 2010). However, prior to this stage in the analysis, between group comparisons were also made and these provided the primary outcomes for the study. The impact of this bias is judged to be unclear as while a between group comparison was made, the long-term maintenance of intervention gains is disguised by the combination of data sets at the final time point.

Schertz 2013 adopted a variable intervention period which could have weighted findings towards a positive conclusion regarding intervention efficacy. Their strategy was to recruit participants who were demonstrably lacking the target joint attention skills. Participants were then paired and randomly assigned to intervention or treatment-as-usual control groups. Intervention then proceeded until an individual participating child had achieved the target skill. At this point, exit assessment measures were taken for that child and their matched pair. Therefore, within each pair, participants experienced the same interval between baseline and post-intervention assessment. However, when analysed as a group, this system ensured that every child in the intervention group had shown significant gains in the target skill, thus biasing the study towards a positive conclusion.

No other potential sources of bias were identified in the studies selected for inclusion in this review. Also, the authors note some examples of particularly good practice in the prevention of bias, including close measurement of treatment adherence in therapist-led (Baghdadli 2013) and parent-training studies (Schertz 2013).

Effects of interventions

See: [Summary of findings for the main comparison](#)

Studies in this review used a wide variety of outcome measures, often using measures specific to their intervention target and sometimes designed specifically for that study. In addition, effect sizes, mean differences reported as standardised mean differences (SMD) and confidence intervals (CI) were not always reported, though we include these below where available. Intervention measures are listed in additional [Table 2](#) and discussed below, organised by primary and secondary outcome category. In addition, primary outcome results are collated in the [Summary of findings for the main comparison](#).

Evaluating primary outcomes: communication

Primary outcome measures in this section were those using standardised assessments to assess communication skills of diagnostic relevance (i.e. more than just expressive language).

Two studies employed diagnostic assessment measures to evaluate change in symptom level within the communication domain.

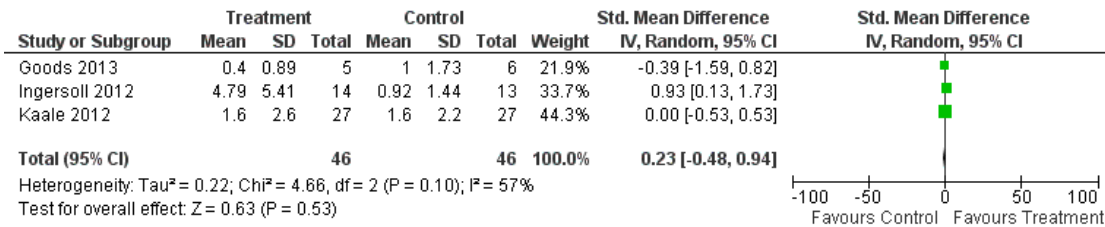
Wong 2010 used a sub-set of ADOS (Lord 2000) items to evaluate communication gains in response to intervention, finding improvements in relevant items (vocalisation directed to others, gestures, pointing) in the intervention group (median difference = 4 points), but not in the control group (median difference = 2.5 points). This finding is weakened by the fact that these analyses compared change from baseline to outcome in each group separately and there was no between-group comparison. Furthermore, the ADOS is not intended as an intervention outcome measure, and it is not usual to analyse a sub-set of items. On the other hand, this finding is strengthened by a comparison that shows no intervention group gains in items pre-identified as non-relevant to the intervention. Young 2012 similarly reported change for individual items of the Social Communication Questionnaire (SCQ) (Rutter 2003). Of specific relevance in this outcome category, they analysed change in eye contact and gaze aversion, and found no intervention effects on these items (effect sizes: $\eta_p^2 = 0.001$ and $\eta_p^2 = 0.002$ respectively).

For participants at a higher level of communicative sophistication, Hadwin 1996 (reported in Hadwin 1997) evaluated the impact of ToM intervention on complex language skills. They found no effect of intervention on conversational skills, and also raised additional evidence that language level may moderate intervention effects when teaching ToM skills (Hadwin 1996) though this is not explored elsewhere.

Evaluating primary outcomes: social function

Primary outcome measures in this section were those using standardised assessments to assess social skills of diagnostic relevance. Six studies used the Early Social Communication Scales (ESCS) to evaluate the outcome, standardised observational assessment of social engagement behaviours, including joint attention (Kasari 2006; Kim 2009; Ingersoll 2012; Kaale 2012; Goods 2013; Wong 2013). Due to differences in measurement (e.g. reporting behaviour frequency versus amount of time; reporting sub-items from the scales versus reporting scale totals) and study design (those with or without a therapeutic control condition), not all studies could be combined for meta-analysis. However, three studies were combined in this way (Analysis 1.3) and the outcome indicated no significant intervention effect on social behaviour measured in this way (SMD 0.23, 95% CI -0.48 to 0.94, $Z = 0.63$, P value = 0.53, three studies, 92 participants). There was no evidence of heterogeneity in effects ($I^2 = 57\%$, Chi^2 ($df = 2$) = 4.66, P value = 0.10, $\text{Tau}^2 = 0.22$) (Figure 4). Among studies not included in meta-analysis, Kasari 2006 found large treatment gains in the showing (effect size = 1.50) and responding to joint attention (effect size = 1.20) items of the ESCS but not in other relevant items such as pointing and giving. This may indicate a lack of generalisation of skills beyond the specific taught items.

Figure 4. Forest plot of comparison: I test, outcome: I.3 Joint attention initiations in standardised assessment.



As described above, two studies used scores on a diagnostic or screening measure as their primary outcome, principally focusing on change in social function. Wong 2010 reported item-by-item changes rather than algorithm totals for the ADOS (Lord 2000), and Young 2012 analysed intervention effects on a single 'Social Peer Interest' item of the SCQ (Rutter 2003). Young 2012 did not find significant effects on this SCQ item ($\eta_p^2 = 0.06$), despite the intervention having a positive impact on emotion recognition skill; evidence of a lack of generalisation or expansion of a taught skill to related domains. In contrast, Wong 2010 did report intervention effects on items from the ADOS, which are relevant to the intervention content (intervention group median difference = 7 points; control group median difference = 2.5 points), and crucially not for items unrelated to the intervention. However, it should once again be noted in both studies that the measures used were not designed to be analysed on an item-by-item basis. Moreover, Wong 2010 reported no direct comparison of the degree of change between the two groups.

Williams 2012 additionally reported on social abilities measured in a real-world context using standardised measures of life skills (intervention group mean difference = 2.13 points; control group mean difference = 1.59 points). No evidence was found of intervention effects measured in this way, once more indicating a lack of generalisation from specifically-taught skills to wider social abilities.

Further detail on outcomes from observed measures of social engagement is provided below under the heading of *Secondary Outcomes: Change in participant behaviour or quality of interpersonal interaction, or both, measured by direct observation.*

Evaluating secondary outcomes

Intervention Specific: Change in targeted cognitive skill, Emotion recognition

Seven studies examined the impact of intervention on facial affect recognition skills from photographs, as compared to treatment-as-usual (Solomon 2004; Golan 2006; Ryan 2010; Hopkins 2011;

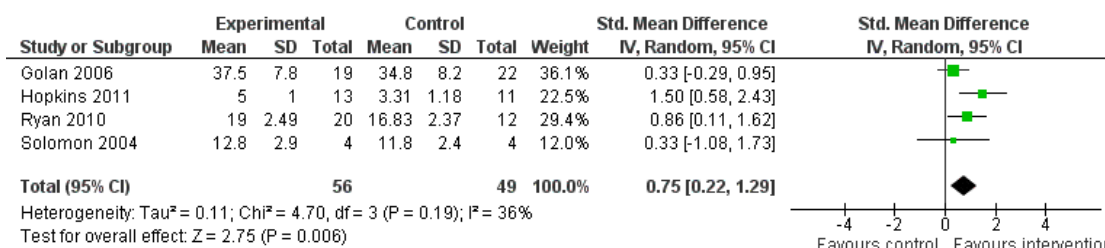
Williams 2012; Young 2012; Baghdadli 2013), which were inspected further for potential inclusion in a meta-analysis. Young 2012 was excluded from the meta-analysis due to use of a non-random allocation procedure. Baghdadli 2013 was eligible for inclusion in a meta-analysis in principle, but it was not possible to extract the relevant data from the paper, which reported median scores and inter-quartile range for its non-normally distributed data, rather than means and standard deviations. The study found a significant improvement in the intervention group (but not the control group) in recognition of angry emotions only (effect size, Cohen's $d = -0.8$, P value = 0.05), but no significant differences between groups for other emotions. It was also inappropriate to incorporate Williams 2012 into the meta-analysis as this study used a control condition, which was hypothesised to have potential treatment effects.

Hopkins 2011 reported on two separate samples: children with low-functioning autism (LFA) and children with high-functioning autism (HFA). All other studies in this group reported on participant samples with IQ or language ability in the normal range, and therefore it seemed most appropriate to include the HFA sample from Hopkins 2011 in the meta-analysis.

Likewise, Solomon 2004 reported separately on younger and older groups of children. The majority of other studies in this group reported on child participants whose age more closely matches the older group of Solomon et al (Ryan 2010; Hopkins 2011), and one remaining study involved an adult sample (Golan 2006). Therefore, it was decided to include the older participant group in the meta-analysis.

All studies in the emotion recognition group reported significant group differences on outcome immediately post-treatment, measured by recognition of facial emotion from static images. The meta-analysis (Analysis 1.2) shows evidence of a positive intervention effect on emotion recognition (SMD 0.75, 95% CI 0.22 to 1.29, Z = 2.75, P value = 0.006, four studies, 105 participants). There was no evidence of heterogeneity in effects ($I^2 = 36%$, Chi² (df = 3) = 4.70, P value = 0.19, Tau² = 0.11) (Figure 5). A study excluded for reasons of bias (Young 2012) also found a positive effect of intervention on emotion recognition skills.

Figure 5. Forest plot of comparison: I test, outcome: I.2 Emotion recognition from face photographs, TAU control.



We note that [Bolte 2002](#) reported a null finding in a related study using emotion recognition training, with recognition of facial emotion from static images as an outcome. However, it was not possible to retrieve these data. This study had a small sample size (n = five per group) and therefore it is unlikely that this null finding would have a large effect on the reported meta-analysis. In addition to the results included in our meta-analyses, many interventions targeting emotion recognition skills were also evaluated by additional measures tapping the same or related constructs. For example, emotion recognition was assessed using stimuli, including audio clips ([Golan 2006](#)), film clips ([Golan 2006](#)), emotional vocabulary tests ([Golan 2010](#)), and matching emotional vignettes to facial expressions ([Golan 2010](#)). A consistent finding in these studies using multiple measures of the same construct was an absence of generalisation of the target skill to novel settings or stimuli where this was assessed. For example, [Golan 2006](#) evaluated emotion recognition and found improvements only on 'close-generalisation' tasks, which did not extend to other outcome measures (e.g. reading-the-mind-in-the-eyes: intervention group mean difference = + 0.7 points; control group mean difference = -0.9 points).

Intervention Specific: Change in targeted cognitive skill, Theory of Mind (ToM)

Four studies, all aiming to directly teach ToM, used explicit assessment of ToM as their primary outcome ([Hadwin 1996](#); [Solomon 2004](#); [Fisher 2005](#); [Begeer 2011](#)). These outcome measures overlap with those described in the *Emotion Recognition* section above as they sometimes use emotional content (especially complex emotions and mental states). However, they are evaluated independently here as the source authors clearly identify them as assessments of ToM. Once again, positive outcomes were found when assessing progress within, or close to, the taught context, but there was an absence of generalisation of taught skills to novel, or more complex scenarios or to abilities hypothetically built on ToM. For example, [Begeer 2011](#) reported a significant difference between groups in degree of improvement measured by the ToM test (effect size, Cohen's d = 0.75), but no such effect in measures of self-

reported empathy or parent-reported social skills.

[Williams 2012](#) reported on an emotion skills intervention but additionally assessed ToM abilities in order to explore the extent of skills learnt. Once more, there were no intervention effects on this extended skill set.

Intervention Specific: Change in targeted cognitive skill, Imitation

Two studies reported gains in imitation skills as their primary outcome in intervention studies with toddlers ([Landa 2011](#); [Ingersoll 2012](#)). For example, [Ingersoll 2012](#) (reported in [Ingersoll 2010](#)) showed that the treatment group made larger gains in imitation than the control group, though this finding had a small to moderate effect size (elicited imitation, $\eta_p^2 = 0.20$; spontaneous imitation, $\eta_p^2 = 0.29$; object imitation $\eta_p^2 = 0.21$; gesture imitation $\eta_p^2 = 0.38$). Unfortunately, due to differences in measurement it was not possible to combine these two studies in meta-analysis. Their combined sample size is just 69 participants.

Intervention Specific: Change in targeted cognitive skill, Play

Finally, four studies included assessment of play as a secondary outcome ([Hadwin 1996](#), [Wong 2010](#); [Goods 2013](#); [Wong 2013](#)). [Hadwin 1996](#) found no effect of teaching ToM understanding on observed symbolic play skills, while the findings of [Wong 2010](#), using an observational symbolic play test are positive, though the larger sample size is reported by [Hadwin 1996](#). Regarding the range of play types observed, [Goods 2013](#) reported positive intervention effects on this variable using a 'Structured Play Assessment' while [Wong 2013](#) reported no positive effects on the same measure. These two studies each reported a different output from the 'Structured Play Assessment' (play types versus play level) making it illogical to combine these data in a meta-analysis.

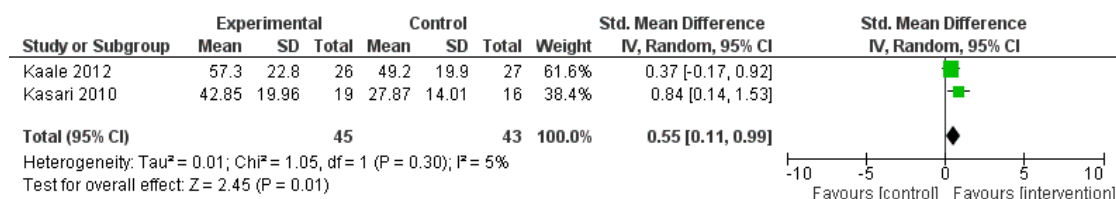
Change in participant behaviour or quality of interpersonal interaction, or both, measured by direct observation

A series of studies used joint engagement behaviours during adult-child interactions as an outcome measure. However, subtle but important differences in the operationalisation of these outcomes meant that very few were eligible for combined analysis. For example, some studies reported on parent-child (Kasari 2006; Kasari 2010) and some on teacher-child interactions (Kim 2009; Kaale 2012; Goods 2013). Some studies reported on only one category of joint attention behaviour as a single outcome (Kaale 2012), while some sub-divided these into initiations and responses (Kasari 2006; Kasari 2010; Landa 2011). In addition, there were differences in the outcome unit of measurement with Kasari 2006 reporting total seconds of joint attention behaviours and Kasari 2010 reporting frequency of joint attention behaviours. And fi-

nally, some studies reported on behaviour measured in a one-to-one setting (Kaale 2012) while others reported on spontaneous behaviour within a group setting, e.g. the classroom (Wong 2013) or during school recess (Hopkins 2011).

The only outcomes of sufficient similarity to be appropriate for inclusion in meta-analysis were the measures of joint engagement during mother-child play reported in Kasari 2010 and Kaale 2012, both as percentages of total time. In meta-analysis (Analysis 1.1) there was evidence of a positive intervention effect on joint engagement (SMD 0.55, 95% CI 0.11 to 0.99, $Z = 2.45$, P value = 0.01, two studies, 88 participants). There was no evidence of heterogeneity in effects ($I^2 = 5\%$, Chi^2 ($df = 1$) = 1.05, P value = 0.30, $\text{Tau}^2 = 0.01$) (Figure 6).

Figure 6. Forest plot of comparison: 1 test, outcome: 1.1 Joint engagement in mother-child interaction.



Here, there was a little evidence that taught skills may generalise to new settings. Kasari 2006 and Kaale 2012 both found that therapist-taught social communication skills were in evidence in a parent-child interaction scenario and when using novel assessment materials (Gulsrud 2007). For example, Kaale 2012 found that children in the intervention group spent on average 12.2% more time in a joint engagement state with their mother compared with children in the control group (95% CI = 2.4% to 22%, effect size Cohen's $d = 0.67$).

Change in participant behaviour or skills measured by adult report

Other studies reported an array of general social skills measures (Solomon 2004; Fisher 2005; Kim 2009; Wong 2010; Begeer 2011; Hopkins 2011; Ingersoll 2012). These were very different in method of assessment, construct being examined, and data format, and so it is not possible to make a direct comparison between studies. Nevertheless, the overall message from the study authors is of improvement in social skills as a result of intervention.

Change in participant cognitive skill, measured by standardised assessment

Three studies reported measures of language and general cognitive or adaptive ability as an outcome (Kasari 2006 reported in Kasari 2008, Landa 2011; Schertz 2013). The studies produced conflicting results. Landa 2011 did not find a significant difference between intervention and control groups on a measure of expressive language despite a moderate effect size of 0.49, while Kasari 2008 and Schertz 2013 reported significant gains in expressive language in the intervention compared with the control group. However, when evaluating participants over the long term (Kasari 2012a), there was no evidence that treatment continued to impact on language and cognitive outcome five years from baseline.

Acceptability of Intervention

A range of studies incorporated caregiver measures such as quality of involvement, adherence to treatment, mental health or participant well-being and satisfaction surveys (Solomon 2004; Kasari 2010; Wong 2010; Baghdadi 2013; Wong 2013), but these were very different in method of assessment and specific construct targeted, and so it is not possible to draw solid conclusions from these data.

Rate of drop-out

As discussed previously ([Assessment of risk of bias in included studies](#)), rates of drop-out in the included studies were very low with half of the included studies reporting that the full sample remained enrolled for the length of the trial. One study reported a very high rate of drop-out from a small sample in a study taking place within a pre-school setting and it was unclear why this drop-out rate was so high ([Goods 2013](#)). Otherwise, the highest rates of drop-out occurred when the intervention was self-directed by an individual with ASD ([Golan 2006](#)) or when the intervention took place at a clinic which required significant travelling time for participants ([Kim 2009](#)).

Economic data

No studies reported economic data.

Follow-up effects

A minority of studies followed up participants after the immediate intervention period had ended ([Hadwin 1996](#); [Fisher 2005](#); [Kasari 2006](#); [Kasari 2010](#); [Landa 2011](#); [Ryan 2010](#)). Follow-up periods ranged from six weeks to five years and findings consistently showed maintenance of treatment gains in a range of measured skills, including ToM ([Hadwin 1996](#); [Fisher 2005](#)), social engagement ([Kasari 2006](#); [Kasari 2010](#); [Landa 2011](#)), and emotion recognition ([Ryan 2010](#)). In some cases there was evidence of continued significant growth during the post-intervention period ([Fisher 2005](#); [Kasari 2006](#); [Kasari 2010](#)). The sole exception is [Kasari 2006](#) (reported in [Kasari 2012a](#)) who found no long-term impact of intervention on language and cognitive outcome at five years post-baseline.

DISCUSSION

Summary of main results

Twenty-two randomised trials were included in this review. All reported on interventions that targeted either Theory of Mind (ToM), or one of the accepted sub-components of ToM such as shared attention, emotion recognition or imitation. Most involved either wait-list or treatment-as-usual control conditions, or a control condition devised to replicate contact time but without therapeutic content.

Risk of bias in the included studies was variable. Very few studies provided adequate information and in a majority of cases the authors were contacted to provide further details. In particular, details of sequence generation and allocation concealment were lacking in the published articles. Blinding of participants and personnel was judged to be impossible in almost every study reviewed, and as a result risk of bias in this category was high. Blinding of

outcome assessors is better achieved and reported on, and concerns about selective reporting or attrition were rare.

In addition, outcomes varied widely, both in the construct being measured and the means of measurement. This made it a challenge to combine studies for meta-analysis and as a result only three, very small meta-analyses were conducted. Outcomes also differed significantly from those predicted in the review protocol. There were very few outcomes reported that fell into the protocol-defined categories of symptom level for the two diagnostic domains of social and communication impairments. Instead, studies mostly reported on intervention-specific measures of constructs such as emotion recognition, joint attention and social communication, and imitation. In these three intervention target areas (emotion recognition, joint attention and social communication, and imitation), there was modest evidence of intervention success.

Emotion recognition

Studies in the emotion recognition field consistently found positive effects of specific training, a conclusion reinforced by combined analysis of four studies ([Figure 5](#)). However these skills did not always generalise to testing contexts that moved beyond the trained content. Since the overall goal of emotion recognition training for people with ASD must be to improve real world emotion recognition and, by extension, emotional understanding, this is a significant limitation, which calls into question the value of emotion recognition interventions. These studies are additionally limited by their reliance on both teaching and measuring facial affect recognition as a proxy for emotion recognition. Real world emotion recognition skills require appraisal of emotional valence from a variety of sources (e.g. facial emotion, body language, content and tone of speech, context) and these skills are rarely taught.

Joint attention and social communication

A number of included studies reported on therapist-led, one-to-one or group interventions, normally with young children with autism, targeting a core deficit of joint attention and social communication. These studies often involved a high level of intervention contact hours but could produce sustained effects ([Kasari 2010](#)), although the intervention effect may not be maintained in the longer term, i.e. over a period of years ([Kasari 2006](#)). The studies in this category have significant real world relevance as there is evidence that therapist training can lead to improvements in interactions with other familiar adults as well ([Kasari 2006](#); [Kaale 2012](#)). Since joint attention is a theoretical prerequisite for both language learning and ToM development, more longitudinal studies looking at the effect of joint attention improvement on these linked skills would be of value.

Imitation

One study specifically targeted imitation both as intervention target and method of delivery (Ingersoll 2012), while another used a more general social communicative approach which impacted on imitation skills (Landa 2011). These studies, in common with those described above, require further investigation to explore whether training on this theoretical building block of ToM can subsequently enhance ToM development.

Theory of Mind

Perhaps surprisingly, a minority of included studies focused explicitly on training ToM skills. As for emotion recognition, it was found relatively straightforward to teach a core ToM skill and to find a difference in that skill when tested under the same circumstances, but generalisation across props, settings, and partners was a greater challenge.

General conclusions

A common theme, therefore, in these studies was the difficulty of creating positive intervention effects recognisable in everyday function. The skills targeted in the interventions included in this review are all developmentally linked, both to each other and to more general abilities such as language and friendship skills. However, it is rare for an intervention to succeed in triggering this developmental chain. For this reason it may be that the broad-based interventions, excluded here because they cannot be clearly categorised according to the protocol criteria, represent a more successful approach to intervention for autism. They may help to promote generalisation by working in a more ecologically valid way to support the development of new skills in combination.

One complicating factor, which hinders the opportunity to draw strong conclusions from the review, is that the studies lack appropriate outcome measures in two different ways. First, there is no consistent, widely-used measure for assessing intervention outcome using diagnosis-relevant domains for ASD research - not even for a specific age-range or ability level. As a result, studies often rely on bespoke measures such as video-coding of parent-child interactions, or multiple measures assessing a range of social, intellectual, and adaptive skills. This is an obstacle to comparison of multiple studies both at a systematic review level and for the average reader of these published reports. Second, the use of multiple measures, combined with an absence of explicit identification of a single primary outcome measure, means that it is often possible to claim intervention success on a sub-set of the measures used, but not all. Without a priori identification of the primary outcome measure it is difficult to judge how these mixed findings should be interpreted.

Overall completeness and applicability of evidence

The studies reported in this review cover a good range of ToM and associated skills, though the fact of this array means that more evidence is needed in each category. It was not possible to conduct subgroup analyses for this review, but it is theoretically likely that we would find differences between specific diagnostic categories, age groups, and intervention delivery methods. The evidence, though limited and of generally low quality, suggests that it may be possible to teach ToM-linked abilities.

Research now should explore how different approaches might be applied to different populations, requiring a much larger and more systematic body of evidence. In particular, it is understood that different age and ability levels require support developing different sub-skills and mapping out these relationships would be of value. In the field of emotion recognition, however, this skill has been successfully taught to preschoolers, children and adults, quite often using computer-based methods, and an understanding of why this is a difficulty which persists across the life span would be of value.

Quality of the evidence

The evidence included here varies in quality due to the aforementioned difficulty of blinding participants and personnel, and under-reporting of other relevant practices such as sequence generation and allocation concealment. In addition, sample sizes are often very low, with the maximum reported at only $n = 61$ (Kaale 2012). Due to a lack of strong reporting norms it has not always been possible to accurately judge risk of bias in the studies reported here, which may have unknown impact on conclusions. Further, while attrition rates are low on the whole, all studies analyse only the final sample and do not use an intention-to-treat analysis. Once more it is impossible to judge to what degree this may have affected results, and in what direction. Overall, there is a lack of good quality evidence in this field and a requirement for more randomised controlled trials representing highest standards in methodology, particularly outcome assessment.

One group of studies represented a good example of independent evaluation of the same intervention, *The Transporters* DVD for teaching emotion recognition and understanding (Golan 2010; Williams 2012; Young 2012). Unfortunately, minor but important differences in study design between these reports meant that they could not be combined in a meta-analysis. However this rare example of replication, perhaps because the DVD-based intervention is relatively easy to apply, is a notable and positive exception in a field of diverse intervention strategies.

We found no evidence of adverse effects but this partly reflects a lack of attention to these in the source literature. There seems to be a common presumption that behavioural interventions, such as those reviewed here, do not carry a risk of harm but this possibility needs to be addressed explicitly in future studies.

Potential biases in the review process

The review authors were restricted by only evaluating studies available in English language. In addition, some databases searched at the first data collection point were no longer available to us when searches were repeated two years later. It is unclear how these restrictions may have affected the conclusions drawn. Dissertation databases were not searched independently and instead the review authors requested unpublished data, including dissertations, from key authors. Two relevant PhD theses were examined (Arora 2008; Rodgers 2012) one of which is an excluded study (Rodgers 2012). We note that it is extremely rare in this field for postgraduate dissertations to conduct an independent clinical trial because of the challenges surrounding recruitment and design.

Agreements and disagreements with other studies or reviews

As this review appears to be the first to consider ToM interventions as a group of studies, no comparison can be made with the conclusions of other reviews.

AUTHORS' CONCLUSIONS

Implications for practice

The review suggests it may be possible to teach both Theory of Mind (ToM) and the precursor skills associated with the construct. However this teaching rarely or never generalises to novel contexts, and it is unclear whether there is long-term maintenance of learnt skills, or developmental progression in learning. Future reviewers should consider investigating the results of multi-modal interventions, which were excluded from this review, and which teach a range of real-world social and communicative skills as well as academic learning. Given the problems with maintenance and generalisation of taught skills, multi-modal approaches may be more successful in improving social and communication outcomes in people with ASD.

Implications for research

The ToM model of autism development proposes that ToM is a core deficit, which links both to precursor skills, such as joint attention and emotion recognition, and to subsequent abilities such as making friends and understanding metaphorical language. Theoretically then, this developmental model implies that a successful intervention at one point in the chain should have impact all the way up the developmental ladder.

More longitudinal studies are required to establish whether teaching a specific skill at one time point will lead not only to maintenance

of that skill but also the development of further associated abilities subsequently. Studies systematically linking intervention targets and delivery methods to participant age and ability would support this process by helping to identify the right 'starting place' for an intervention for a particular individual. In order to reach larger populations, multi-site studies are required, with co-operation across research centres and national borders. This is challenging not just in a logistical sense but also because countries and even counties vary in the service provision and intervention approaches offered to people with ASD across ages.

There is an urgent need for better outcome measures for autism research across the board. This has been recognised in current funding from the National Institute for Health Research Health Technology Assessment programme for a systematic review of measurement properties of outcome measures for children with autism up to the age of six years (<http://www.hta.ac.uk/project/2830.asp>). For future studies in the field of ToM interventions, ideally outcome measures should capture symptom severity in core diagnostic domains, without looking for a change in diagnostic status - which is neither feasible nor necessarily to be desired. In the short term, however, the outcome measure issue could be partially resolved by better study reporting, in particular a priori identification of a single primary outcome measure to define intervention success. Additionally, secondary measures should have a clear rationale for inclusion and should be, wherever possible, linked to intervention components. Improvements in reporting could also be made, especially in providing details of sequence generation and allocation concealment, which is rarely described.

The Theory of Mind model of autism was first introduced in 1985 and in 2000 a review of the changes in the model was published (Baron-Cohen 2000). Now, almost another 15 years on, multiple versions (Astington 2011) and alternative explanatory models abound (e.g. Happe 2006; Motttron 2006; Baron-Cohen 2010) and there have been direct attacks both on the model (Hobson 1991) and on the methods used to measure the construct (Bloom 2000). One meaningful way to evaluate the explanatory power and clinical and educational relevance of these competing theories is to explore their impact in an intervention setting. Currently, however, the quality and quantity of evidence needed even for ToM alone is inadequate to do so.

ACKNOWLEDGEMENTS

The authors are grateful for the support of Jo Abbott, Margaret Anderson, Chris Champion, Jane Dennis, Laura MacDonald, Geraldine MacDonald, and Inalegwu Oono for their help in developing this review and for the comments of Michelle Dawson on a previous draft.

REFERENCES

References to studies included in this review

Baghdadli 2013 *{published data only}*

Baghdadli A, Brisot J, Henry V, Michelon C, Soussana M, Rattaz C, et al. Social skills improvement in children with high-functioning autism: a pilot randomized controlled trial. *European Child & Adolescent Psychiatry* 2013;**22**(7): 433–42.

Begeer 2011 *{published data only}*

Begeer S, Gevers C, Clifford P, Verhoeve M, Kat K, Hoddenbach E, et al. Theory of mind training in children with autism: a randomized controlled trial. *Journal of Autism and Developmental Disorders* 2011;**41**(8):997–1006.

Bolte 2002 *{published data only}*

* Bolte S, Feineis-Matthews S, Leber S, Dierks T, Hubl D, Poutska F. The development and evaluation of a computer-based program to test and teach the recognition of facial affect. *International Journal of Circumpolar Health* 2002;**61 Suppl 2**:61–8.

Bolte S, Hubl D, Feineis-Matthews S, Prvulovia D, Dierks T, Poutska F. Facial affect recognition training in autism: can we animate the fusiform gyrus?. *Behavioural Neuroscience* 2006;**120**(1):211–6.

Fisher 2005 *{published data only}*

Fisher N, Happe F. A training study of theory of mind and executive function in children with autistic spectrum disorders. *Journal of Autism and Developmental Disorders* 2005;**35**(6):757–71.

Golan 2006 *{published data only}*

Golan O, Baron-Cohen S. Systemizing empathy: teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia. *Development and Psychopathology* 2006;**18**(2): 591–617.

Golan 2010 *{published data only}*

Golan O, Ashwin E, Granader Y, McClintock S, Day K, Leggett V, et al. Enhancing emotion recognition in children with autism spectrum conditions: an intervention using animated vehicles with real emotional faces. *Journal of Autism and Developmental Disorders* 2010;**40**(3):269–79.

Goods 2013 *{published data only}*

Goods KS, Ishijim E, Chang Y-C, Kasari C. Preschool based JASPER intervention in minimally verbal children with autism: pilot RCT. *Journal of Autism and Developmental Disorders* 2013;**43**(5):1050–6.

Hadwin 1996 *{published data only}*

* Hadwin J, Baron-Cohen S, Howlin P, Hill K. Can we teach children with autism to understand emotions, belief or pretence?. *Development and Psychopathology* 1996;**8**(2): 345–65.

Hadwin J, Baron-Cohen S, Howlin P, Hill K. Does teaching theory of mind have an effect on the ability to develop conversation in children with autism?. *Journal of Autism and Developmental Disorders* 1997;**27**(5):519–37.

Hopkins 2011 *{published data only}*

Hopkins IM, Gower MW, Perez TA, Smith DS, Amthor FR, Wimsatt FC, et al. Avatar assistant: improving social skills in students with an ASD through a computer-based intervention. *Journal of Autism and Developmental Disorders* 2011;**41**(11):1543–55.

Ingersoll 2012 *{published and unpublished data}*

* Ingersoll B. Brief report: effect of a focused imitation intervention of social functioning in children with autism. *Journal of Autism and Developmental Disorders* 2012;**42**(8): 1768–73.

Ingersoll B. Brief report: pilot randomized controlled trial of reciprocal imitation training for teaching elicited and spontaneous imitation to children with autism. *Journal of Autism and Developmental Disorders* 2010;**40**(9):1154–60. [PUBMED: 20155309]

Kaale 2012 *{published data only}*

Kaale A, Smith L, Sponheim E. A randomized controlled trial of preschool-based joint attention intervention for children with autism. *Journal of Child Psychology and Psychiatry* 2012;**53**(1):97–105.

Kasari 2006 *{published data only}*

Arora T. *Perseveration in young children with autism and the association with joint attention [PhD Thesis]*. University of California, 2008.

Gulsrud AC, Kasari C, Freeman S, Paparella T. Brief report: children with autism's response to novel stimuli while participating in interventions targeting joint attention or symbolic play skills. *Autism* 2007;**11**(6):535–46.

* Kasari C, Freeman S, Paparella T. Joint attention and symbolic play in young children with autism: a randomized controlled intervention study. *Journal of Child Psychology and Psychiatry* 2006;**47**(6):611–20.

Kasari C, Gulsrud A, Freeman S, Paparella T, Helleman G. Longitudinal follow up of children with autism receiving targeted interventions on joint attention and play. *Journal of the American Academy of Child and Adolescent Psychiatry* 2012;**51**(5):487–95.

Kasari C, Paparella T, Freeman S, Jahromi LB. Language outcome in autism: a randomised comparison of joint attention and play interventions. *Journal of Consulting and Clinical Psychology* 2008;**76**(1):125–37.

Kasari 2010 *{published data only}*

Kasari C, Gulsrud A, Wong C, Kwon S, Locke J. Randomized controlled caregiver mediated joint engagement intervention for toddlers with autism. *Journal of Autism and Developmental Disorders* 2010;**40**(9):1045–56.

Kim 2009 *{published data only}*

* Kim J, Wigram T, Gold C. Emotional, motivational and interpersonal responsiveness of children with autism in improvisational music therapy. *Autism* 2009;**13**(4): 389–409.

Kim J, Wigram T, Gold C. The effects of improvisational music therapy on joint attention behaviours in autistic

children: a randomized controlled study. *Journal of Autism and Developmental Disorders* 2008;**38**(9):1758–66.

Landa 2011 {published data only}

Landa RJ, Holman KC, O'Neill AH, Stuart EA. Intervention targeting development of socially synchronous engagement in toddlers with autism spectrum disorder: a randomized controlled trial. *Journal of Child Psychiatry and Psychology* 2011;**52**(1):13–21.

Ryan 2010 {published and unpublished data}

Ryan C, Charragain CN. Teaching emotions recognition skills to children with autism. *Journal of Autism and Developmental Disorders* 2010;**40**(12):1505–11.

Schertz 2013 {published data only}

Schertz HH, Odom SL, Baggett KM, Sideris JH. Effects of joint attention mediated learning for toddlers with autism spectrum disorders: an initial randomized controlled study. *Early Childhood Research Quarterly* 2013;**28**(2):249–58.

Solomon 2004 {published data only}

Solomon M, Goodlin-Jones BL, Anders TF. A social adjustment enhancement intervention for high functioning autism, Asperger's syndrome and pervasive developmental disorder NOS. *Journal of Autism and Developmental Disorders* 2004;**34**(6):649–68.

Williams 2012 {published data only}

Williams BT, Gray KM, Tonge BJ. Teaching emotion recognition skills to young children with autism: a randomized controlled trial of an emotion training programme. *Journal of Child Psychology and Psychiatry* 2012;**53**(12):1268–76.

Wong 2010 {published data only}

Wong VCN, Kwan QK. Randomized controlled trial for early intervention for autism: a pilot study of the Autism 1-2-3 Project. *Journal of Autism and Developmental Disorders* 2010;**40**(6):677–88.

Wong 2013 {published data only}

Wong CS. A play and joint attention intervention for teachers of young children with autism: a randomized controlled pilot study. *Autism* 2013;**17**:340–57.

Young 2012 {published data only}

Young RL, Posselt M. Using The Transporters DVD as a learning tool for children with autism spectrum disorders. *Journal of Autism and Developmental Disorders* 2012;**42**(6):984–91.

References to studies excluded from this review

Adams 2012 {published data only}

Adams C, Lockton E, Freed J, Gaile J, Earl G, McBean K, et al. The Social Communication Intervention Project: a randomized controlled trial of the effectiveness of speech and language therapy for school-age children who have pragmatic and social communication problems with or without autism spectrum disorder. *International Journal of Language & Communication Disorders* 2012;**47**(3):233–44.

Aldred 2004 {published data only}

Aldred C, Green J, Adams C. A new social communication intervention for children with autism: a pilot randomised

controlled treatment study suggesting effectiveness. *Journal of Child Psychology and Psychiatry* 2004;**45**(8):1420–30. [DOI: 10.1111/j.1469-7610.2004]

Beaumont 2008 {published data only}

Beaumont R, Sofronoff K. A multi-component social skills intervention for children with Asperger syndrome: the Junior Detective Training Program. *Journal of Child Psychology and Psychiatry* 2008;**49**(7):743–53.

Cardon 2011 {published data only}

Cardon TA, Wilcox MJ. Promoting imitation in young children with autism: a comparison of reciprocal imitation training and video modelling. *Journal of Autism and Developmental Disorders* 2011;**41**(5):654–66.

Carter 2011 {published data only}

Carter AS, Messinger DS, Stone WL, Celimli S, Nahmias AS, Yoder P. A randomized controlled trial of Hanen's 'More Than Words' in toddlers with early autism symptoms. *Journal of Child Psychology and Psychiatry* 2011;**52**(7):741–52.

Casenhiser 2011 {published data only}

Casenhiser DM, Shanker SG, Steiben J. Learning through interaction in children with autism: preliminary data from a social-communication-based intervention. *Autism* 2011;**17**(2):220–41. [DOI: 10.1177/1362361311422052]

Castorina 2011 {published data only}

Castorina LL, Negri LM. The inclusion of siblings in social skills training groups for boys with Asperger syndrome. *Journal of Autism and Developmental Disorders* 2011;**41**(1):73–81.

Charman 2007 {published data only}

Charman T, Hudry K. Interventions targeting joint attention and symbolic play can improve aspects of these skills in young children with autism. *Evidence-Based Mental Health* 2007;**10**(1):12–21.

Corbett 2011 {published data only}

Corbett B, Gunther J, Comins D, Price J, Ryan N, Simon D, et al. Brief report: theatre as therapy for children with autism spectrum disorder. *Journal of Autism and Developmental Disorders* 2011;**41**(4):505–11.

Dawson 2010 {published data only}

Dawson G, Rogers S, Munson J, Smith M, Winter J, Greenson J, et al. Randomized, controlled trial of an intervention for toddlers with autism: the Early Start Denver Model. *Pediatrics* 2010;**125**(1):e17–23.

DeRosier 2011 {published data only}

DeRosier M, Swick D, Davis N, McMillen J, Matthews R. The efficacy of a social skills group intervention for improving social behaviors in children with high functioning autism spectrum disorders. *Journal of Autism and Developmental Disorders* 2011;**41**(8):1033–43.

Drew 2002 {published data only}

Drew A, Baird G, Baron-Cohen S, Cox A, Slonims V, Wheelwright S, et al. A pilot randomised control trial of a parent training intervention for pre-school children with autism: preliminary findings and methodological

- challenges. *European Child & Adolescent Psychiatry* 2002;**11**(6):266–72.
- Estes 2011** *{published data only}*
Estes A, Rivera V, Bryan M, Cali P, Dawson G. Discrepancies between academic achievement and intellectual ability in higher-functioning school-aged children with autism spectrum disorder. *Journal of Autism and Developmental Disorders* 2011;**41**(8):1044–52.
- Field 2001** *{published data only}*
Field T, Field T, Sanders C, Nadel J. Children with autism display more social behaviours after repeated imitation sessions. *Autism* 2001;**5**(3):317–23.
- Frankel 2010a** *{published data only}*
Frankel F, Myatt R, Sugar C, Whitham C, Gorospe CM, Laugeson E. A randomized controlled study of parent-assisted children's friendship training with children having autism spectrum disorders. *Journal of Autism and Developmental Disorders* 2010;**40**(7):827–42.
- Frankel 2010b** *{published data only}*
Frankel F, Myatt R, Sugar C, Whitham C, Gorospe CM, Laugeson E. Erratum: a randomized controlled study of parent-assisted children's friendship training with children having autism spectrum disorders. *Journal of Autism and Developmental Disorders* 2010;**40**(7):843.
- Gantman 2012** *{published data only}*
Gantman A, Kapp S, Orenski K, Laugeson E. Social skills training for young adults with high-functioning autism spectrum disorders: a randomized controlled pilot study. *Journal of Autism and Developmental Disorders* 2012;**42**(6):1094–103.
- Gattino 2011** *{published data only}*
Gattino GS, Dos Santos Riesgo R, Longo D, Leite JCL, Faccini LS. Effects of relational music therapy on communication of children with autism: a randomised controlled study. *Nordic Journal of Music Therapy* 2011;**20**(2):142–54.
- Geretsegger 2012** *{published data only}*
Geretsegger M, Holck U, Gold C. Randomised controlled trial of improvisational music therapy's effectiveness for children with autism spectrum disorders (TIME-A): study protocol. *BMC Pediatrics* 2012;**12**(1):2.
- Girolametto 1988** *{published data only}*
Girolametto LE. Improving the social-conversational skills of developmentally disabled children: an intervention study. *Journal of Speech and Hearing Disorders* 1988;**53**(2):156–67.
- Green 2010** *{published data only}*
Green J, Charman T, McConachie H, Aldred C, Slonims V, Howlin P, et al. Parent-mediated communication-focused treatment in children with autism (PACT): a randomised controlled trial. *Lancet* 2010;**375**(9732):2152–60.
- Hanley-Hochdorfer 2010** *{published data only}*
Hanley-Hochdorfer K, Bray MA, Kehle TJ, Elinoff MJ. Social stories to increase verbal initiation in children with autism and Asperger's disorder. *School Psychology Review* 2010;**39**(3):484–92.
- Heimann 2006** *{published data only}*
Heimann M, Laberg KE, Nordoen B. Imitative interaction increases social interest and elicited imitation in non-verbal children with autism. *Infant and Child Development* 2006;**15**(3):297–309.
- Hillier 2012** *{published data only}*
Hillier A, Greher G, Poto N, Dougherty M. Positive outcomes following participation in a music intervention for adolescents and young adults on the autism spectrum. *Psychology of Music* 2012;**40**(2):201–15.
- Howlin 1997** *{published data only}*
Howlin P. When is a significant change not significant?. *Journal of Autism and Developmental Disorders* 1997;**27**(3):347–8.
- Howlin 2007** *{published data only}*
Howlin P, Gordon RK, Pasco G, Wade A, Charman T. The effectiveness of Picture Exchange Communication System (PECS) training for teachers of children with autism: a pragmatic, group randomised controlled trial. *Child Psychology and Psychiatry* 2007;**48**(5):473–81.
- Jocelyn 1998** *{published data only}*
Jocelyn LJ, Casiro OG, Beattie D, Bow J, Kneisz J. Treatment of children with autism: a randomized controlled trial to evaluate a caregiver-based intervention program in community day-care centers. *Journal of Developmental and Behavioural Pediatrics* 1998;**19**(5):326–34.
- Jones 2010** *{published data only}*
Jones K, Howley M. An investigation into an interaction programme for children on the autism spectrum: outcomes for children, perceptions of schools and a model for training. *Journal of Research in Special Educational Needs* 2010;**10**(2):115–23.
- Kasari 2012a** *{published data only}*
Kasari C, Rotheram-Fuller E, Locke J, Gulsrud A. Making the connection: randomised controlled trial of social skills at school for children with autism spectrum disorders. *Journal of Child Psychology and Psychiatry* 2012;**53**(4):431–9.
- Kern 2011** *{published data only}*
Kern JK, Fletcher CL, Garver CR, Mehta JA, Grannemann BD, Knox KR, et al. Prospective trial of equine-assisted activities in autism spectrum disorder. *Alternative Therapies in Health and Medicine* 2011;**17**(3):14–20.
- Koenig 2010** *{published data only}*
Koenig K, White SW, Pachler M, Lau M, Lewis M, Klin A, et al. Promoting social skill development in children with pervasive developmental disorders: a feasibility and efficacy study. *Journal of Autism and Developmental Disorders* 2010;**40**(10):1209–18.
- Lang 2010** *{published data only}*
Lang R. The Autism 1-2-3 Project, a short-duration and low-intensity intervention that targets basic social and communication skills and parent involvement, shows promise and is worthy of future research. *Evidence-Based Communication Assessment and Intervention* 2010;**4**(4):196–9.

- Laugeson 2012** *{published data only}*
Laugeson E, Frankel F, Gantman A, Dillon AR, Mogil C. Evidence-based social skills training for adolescents with autism spectrum disorders: The UCLA PEERS Program. *Journal of Autism and Developmental Disorders* 2012;**42**(6): 1025–36.
- Lawton 2012** *{published data only}*
Lawton K, Kasari C. Teacher-implemented joint attention intervention: pilot randomised controlled study for preschoolers with autism. *Journal of Consulting and Clinical Psychology* 2012;**80**(4):687–95.
- LeGoff 2004** *{published data only}*
LeGoff DB. Use of LEGO as a therapeutic medium for improving social competence. *Journal of Autism and Developmental Disorders* 2004;**34**(5):557–71.
- Lerner 2012** *{published data only}*
Lerner MD, Mikami AY. A preliminary randomized controlled trial of two social skills interventions for youth with high-functioning autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities* 2012;**27**: 147–57.
- Lopata 2010** *{published data only}*
Lopata C, Thomeer ML, Volker MA, Toomey JA, Nida RE, Lee GK, et al. RCT of a manualized social treatment for high-functioning autism spectrum disorders. *Journal of Autism and Developmental Disorders* 2010;**40**(11): 1297–310.
- Mahoney 2005** *{published data only}*
Mahoney G, Perales F. Relationship-focused early intervention with children with pervasive developmental disorders and other disabilities: a comparative study. *Developmental and Behavioural Pediatrics* 2005;**26**(2): 77–85.
- McConachie 2004** *{published data only}*
McConachie H. Commentary on: joint attention and symbolic play in young children with autism: a randomized controlled intervention study. *Child: Care, Health and Development* 2006;**32**(6):752.
- McGregor 1998** *{published data only}*
McGregor E, Whiten A, Blackburn P. Teaching theory of mind by highlighting intention and illustrating thoughts: a comparison of their effectiveness with 3-year-olds and autistic individuals. *British Journal of Developmental Psychology* 1998;**16**(3):281–300.
- Minshew 2010** *{published data only}*
Minshew N, Eack S. Cognitive enhancement therapy for adults with autism spectrum disorders. NIH Clinical Trials Register. [: NCT00902798]
- Morgan 2010** *{published data only}*
Morgan L, Newton L, McCormick J. Comparing parent-implemented interventions for toddlers with autism spectrum disorders. NIH Clinical Trials Register. [: NCT00760812]
- Nefdt 2010** *{published data only}*
Nefdt, N, Koegel R, Singer G, Gerber M. The use of a self-directed learning program to provide introductory training in pivotal response treatment to parents of children with autism. *Journal of Positive Behaviour Interventions* 2010;**12** (1):23–32.
- Oosterling 2010** *{published data only}*
Oosterling I, Visser J, Swinkels S, Rommelse N, Donders R, Woudenberg T, et al. Randomised controlled trial of the focus parent training for toddlers with autism: 1-year outcome. *Journal of Autism and Developmental Disorders* 2010;**40**(12):1447–58.
- Ozonoff 1995** *{published data only}*
Ozonoff S, Miller JN. Teaching theory of mind: a new approach to social skills training for individuals with autism. *Journal of Autism and Developmental Disorders* 1995;**25**(4): 415–33.
- Pajareya 2011** *{published data only}*
Pajareya K, Nopmaneejumruslers K. A pilot randomised controlled trial of DIR/Floortime parent training intervention for pre-school children with autistic spectrum disorders. *Autism* 2011;**15**(2):1–15.
- Quirnbach 2009** *{published data only}*
Quirnbach LM, Lincoln AJ, Feinberg-Gizzo MJ, Ingersoll BR, Andrews SM. Social stories: mechanisms of effectiveness in increasing game play skills in children diagnosed with autism spectrum disorder using a pretest posttest repeated measures randomized control group design. *Journal of Autism and Developmental Disorders* 2009;**39**(2):299–321.
- Roberts 2011** *{published data only}*
Roberts J, Williams K, Carter M, Evans D, Parmenter T, Silove N, et al. A randomised controlled trial of two early intervention programs for young children with autism: centre-based with parent program and home-based. *Research in Autism Spectrum Disorders* 2011;**5**(4):1553–66.
- Rodgers 2012** *{unpublished data only}*
Rodgers JD. *Efficacy of a Social Intervention on Facial Encoding in Children with High-Functioning Autism Spectrum Disorders [PhD thesis]*. Buffalo, New York: University at Buffalo, The State University of New York, 2012.
- Rogers 2010** *{published data only}*
Rogers L, Hemmeter ML, Wolery M. Using a constant time delay procedure to teach foundational swimming skills to children with autism. *Topics in Early Childhood Special Education* 2010;**30**(2):102–11.
- Schertz 2007** *{published data only}*
Schertz HM, Odom SL. Promoting joint attention in toddlers with autism: a parent-mediated developmental model. *Journal of Autism and Developmental Disorders* 2007;**37**(8):1562–75.
- Schmidt 2012** *{published data only}*
Schmidt C, Stichter JP. The use of peer-mediated interventions to promote the generalization of social competence for adolescents with high-functioning autism and Asperger's syndrome. *Exceptionality* 2012;**20**(2): 94–113.

Silver 2001 *{published data only}*

Silver M, Oakes P. Evaluation of a new computer intervention to teach people with autism of Asperger syndrome to recognise and predict emotions in others. *Autism* 2001;**5**(3):299–316.

Smith 2000 *{published data only}*

Smith T, Buch GA, Gamby TE. Parent-directed, intensive early intervention for children with pervasive developmental disorder. *Research in Developmental Disabilities* 2000;**21**(4): 297–309.

Smith 2004 *{published data only}*

Smith T, Groen AD, Wynn JW. Randomized trial of intensive early intervention for children with pervasive developmental disorder. In: Feldman MA editor(s). *Early Intervention: The Essential Readings*. Blackwell, 2004: 153–81.

Stichter 2001 *{published data only}*

Stichter JP. Functional analysis: the use of analogues in applied settings. *Focus on Autism and Other Developmental Disabilities* 2001;**16**(4):232–9.

Strain 2011 *{published data only}*

Strain PS, Bovey EH. Randomized, controlled trial of the LEAP model of early intervention for young children with autism spectrum disorders. *Topics in Early Childhood Special Education* 2011;**31**(3):133–54.

Swettenham 1996 *{published data only}*

Swettenham J. Can children with autism be taught to understand false belief using computers?. *Journal of Child Psychology and Psychiatry* 1996;**37**(2):157–65.

Tanaka 2010 *{published data only}*

Tanaka JW, Wolf JM, Klaiman C, Koenig K, Cockburn J, Herlihy L, et al. Using computerized games to teach face recognition skills to children with autism spectrum disorder: the Let's Face It! program. *Journal of Child Psychology and Psychiatry* 2010;**51**(8):944–52.

Temple 2007 *{published data only}*

Temple K. A randomized comparison of the effect of two prelinguistic communication interventions on the acquisition of spoken communication in preschoolers with ASD. *Child: Care, Health and Development* 2007;**33**(3): 348–9.

Thomeer 2012 *{published data only}*

Thomeer ML, Lopata C, Volker MA, Toomey JA, Lee GK, Smerback AM, et al. Randomized clinical trial replication of a psychosocial treatment for children with high-functioning autism spectrum disorders. *Psychology in the Schools* 2012; **49**(10):942–54.

Turner-Brown 2008 *{published data only}*

Turner-Brown LM, Perry TD, Dichter G, Bodfish J, Penn DL. Brief report: feasibility of social cognition and interaction training for adults with high-functioning autism. *Journal of Autism and Developmental Disorders* 2008;**38**(9): 1777–84.

Wellman 2002 *{published data only}*

Wellman HM, Baron-Cohen S, Caswell R, Gomez JC, Swettenham J, Toye E, et al. Thought-bubbles help children

with autism acquire an alternative to a theory of mind. *Autism* 2002;**6**(4):343–63.

Whalen 2003 *{published data only}*

Whalen C, Schreibman L. Joint attention training for children with autism using behavior modification procedures. *Journal of Child Psychology and Psychiatry* 2003; **44**(3):456–68.

White 2010 *{published data only}*

White SW, Koenig K, Scahill L. Group social skills instruction for adolescents with high-functioning autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities* 2010;**25**(4):209–19.

Wood 2009 *{published data only}*

Wood JJ, Drahota A, Sze K, Dyke M, Decker K, Fujii C, et al. Brief report: effects of cognitive behavioral therapy on parent-reported autism symptoms in school-age children with high-functioning autism. *Journal of Autism and Developmental Disorders* 2009;**39**(11):1608–12.

Yoder 2006 *{published data only}*

Yoder P, Stone W. A randomised comparison of the effect of two prelinguistic communication interventions on the acquisition of spoken communication in preschoolers with ASD. *Journal of Speech, Language and Hearing Research* 2006;**49**(4):698–711.

Yoder 2006b *{published data only}*

Yoder P, Stone WL. Randomised comparison of two communication interventions for preschoolers with autism spectrum disorders. *Journal of Consulting and Clinical Psychology* 2006;**74**(3):426–35.

References to studies awaiting assessment**Baghdadli 2010** *{published data only}*

Baghdadli A, Bristo-Dubois J, Picot M, Michelon C. Comparison of the effect of two prosocial interventions about the evolution of recognition of facial expression abilities and social cognition of children with an Asperger syndrome or high functioning autism [Comparaison de l'effet de deux interventions prosociales sur l'évolution des capacités d'identification des expressions faciales et du raisonnement social d'enfants avec un syndrome d'Asperger ou autisme de haut niveau]. *Neuropsychiatrie de l'Enfance et de l'Adolescence* 2010;**58**(8):456–62.

Additional references**APA 1994**

APA. *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC: American Psychiatric Association, 1994.

APA 2013

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders - V*. Washington, DC: APA, 2013.

Arora 2008

Arora T. *Perseveration in Young Children with Autism and the Association with Joint Attention [PhD thesis]*. University of California, 2008.

Astington 2011

Astington JW, Gopnik A. Theoretical explanations of children's understanding of the mind. *British Journal of Developmental Psychology* 2011;**9**(1):7–31.

Atladottir 2007

Atladottir HO, Parner ET, Schendel D, Dalsgaard S, Thomsen PH, Thorsen P. Time trends in reported diagnoses of childhood neuropsychiatric disorders. *Archives of Pediatrics & Adolescent Medicine* 2007;**161**(2):193–8.

Baird 2006

Baird G, Simonoff E, Pickles A, Chandler S, Loucas T, Meldrum D, et al. Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). *Lancet* 2006;**368**(9531):210–5.

Baron-Cohen 1985

Baron-Cohen S, Leslie AM, Frith U. Does the autistic child have a “theory of mind”? *Cognition* 1985;**21**(1):37–46.

Baron-Cohen 1995

Baron-Cohen S. *Mindblindness*. London, UK: MIT Press, 1995.

Baron-Cohen 2000

Baron-Cohen S. Theory of mind and autism: a fifteen year review. In: Baron-Cohen S, Tager-Flusberg H, Cohen DJ editor(s). *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience*. Oxford: Oxford University Press, 2000:3–20.

Baron-Cohen 2010

Baron-Cohen S. The empathizing-systematizing (E-S) theory of autism. In: Goswami U editor(s). *The Wiley-Blackwell Handbook of Childhood Cognitive Development*. 2nd Edition. Oxford, UK: Wiley-Blackwell, 2010.

Bloom 2000

Bloom P, German TP. Two reasons to abandon the false belief task as a test of theory of mind. *Cognition* 2000;**77**(1):B25–31.

Bowler 1992

Bowler DM. “Theory of mind” in Asperger's syndrome. *Journal of Child Psychology and Psychiatry* 1992;**35**(5):877–93.

Bowler 1997

Bowler DM. Reaction times to mental state and non-mental state questions in false belief tasks by high-functioning individuals with autism. *European Child and Adolescent Psychiatry* 1997;**6**(3):160–5.

Carruthers 1996

Carruthers P, Smith PK. *Theories of Theories of Mind*. Cambridge, UK: Cambridge University Press, 1996.

Charman 2000

Charman T, Baron-Cohen S, Swettenham J, Baird G, Cox A, Drew A. Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cognitive Development* 2000;**15**(4):481–98.

Croen 2002

Croen LA, Grether JK, Hoogstrate J, Selvin S. The changing prevalence of autism in California. *Journal of Autism and Developmental Disorders* 2002;**32**(3):207–15.

Elsabbagh 2012

Elsabbagh M, Divan G, Koh Y-J, Kim YS, Kauchali S, Marciniak C, et al. Global prevalence of autism and other pervasive developmental disorders. *Autism Research* 2012;**5**(3):160–79.

Fombonne 2001

Fombonne E, Simmons H, Ford T, Meltzer H, Goodman R. Prevalence of pervasive developmental disorders in the British nationwide survey of child mental health. *Journal of the American Academy of Child and Adolescent Psychiatry* 2001;**40**(7):827–9.

Frith 1994

Frith U, Happe F, Siddons F. Autism and theory of mind in everyday life. *Social Development* 1994;**3**(2):108–24.

Garfield 2001

Garfield JL, Peterson CC, Perry T. Social cognition, language acquisition and the development of theory of mind. *Mind and Language* 2001;**16**(5):494–541.

Gulsrud 2007

Gulsrud A, Kasari C, Freeman S, Paparella T. Brief report: children with autism's response to novel stimuli while participating in interventions targeting joint attention or symbolic play skills. *Autism* 2007;**11**(6):535–46.

Hadwin 1997

Hadwin J, Baron-Cohen S, Howlin P, Hill K. Does teaching theory of mind have an effect on the ability to develop conversation in children with autism? *Journal of Autism and Developmental Disorders* 1997;**27**(5):519–37.

Happe 1994

Happe F. Wechsler IQ profile and theory of mind in autism: a research note. *Journal of Child Psychology and Psychiatry* 1994;**35**(8):1461–71.

Happe 1995

Happe F. The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development* 1995;**66**(3):843–55.

Happe 2006

Happe F, Frith U. The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders* 2006;**36**(1):5–25.

Hartley 2009

Hartley SL, Sikora DM. Sex differences in autism spectrum disorder: an examination of developmental functioning, autistic symptoms, and coexisting behavior problems in toddlers. *Journal of Autism and Developmental Disorders* 2009;**39**(12):1715–22.

Higgins 2008

Higgins JPT. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester: John Wiley & Sons, 2008.

Hobson 1991

Hobson RP. Against the theory of 'Theory of mind'. *British Journal of Developmental Psychology* 1991;**9**(1):33–51.

Ingersoll 2010

Ingersoll B. Brief report: pilot randomized controlled trial of reciprocal imitation training for teaching elicited and spontaneous imitation to children with autism. *Journal of Autism and Developmental Disorders* 2010;**40**(9):1154–60. [PUBMED: 20155309]

Joseph 2004

Joseph RM, Tager-Flusberg H. The relationship of theory of mind and executive functions to symptom type and severity in children with autism. *Development and Psychopathology* 2004;**16**(1):137–55.

Kanner 1943

Kanner L. Autistic disturbances of affective contact. *Nervous Child* 1943;**2**:217–50.

Kasari 2008

Kasari C, Paparella T, Freeman S, Jahromi LB. Language outcome in autism: randomized comparison of joint attention and play interventions. *Journal of Consulting and Clinical Psychology* 2008;**76**(1):125–37.

Kasari 2012b

Kasari C, Gulsrud A, Freeman S, Paparella T, Helleman G. Longitudinal follow up of children with autism receiving targeted interventions on joint attention and play. *Journal of the American Academy of Child and Adolescent Psychiatry* 2012;**51**(5):487–95.

Kim 2008

Kim J, Wigram T, Gold C. The effects of improvisational music therapy on joint attention behaviours in autistic children: a randomized controlled study. *Journal of Autism and Developmental Disorders* 2008;**38**(9):1758–66.

Klin 2005

Klin A. Clinical evaluation in autism spectrum disorders: psychological assessment within a trans-disciplinary framework. *Handbook of Autism and Pervasive Developmental Disorders*. Wiley, 2005.

Kogan 2009

Kogan MD, Blumberg SJ, Schieve LA, Boyle CA, Perrin JM, Ghandour RM, et al. Prevalence of parent-reported diagnosis of autism spectrum disorder among children in the US, 2007. *Pediatrics* 2009;**124**(5):1395–403.

Lefebvre 2008

Lefebvre C, Manheimer E, Glaville J. Chapter 6: Searching for studies. In: Higgins JPT, Green S editor(s). *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester: John Wiley & Sons, 2008.

Lord 1994

Lord C, Rutter M, Le Couteur A. 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* 1994;**24**(5):659–85.

Lord 2000

Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore PC, et al. The Autism Diagnostic Observation Schedule - Generic: a standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders* 2000;**30**(3):205–23.

Melzoff 1993

Melzoff AM, Gopnik A. The role of imitation in understanding persons and developing a theory of mind. In: Baron-Cohen S, Tager-Flusberg H, Cohen DJ editor(s). *Understanding Other Minds: Perspectives from Autism*. 1st Edition. Oxford, UK: Oxford University Press, 1993.

Mottron 2006

Mottron L, Dawson M, Soulières I, Hubert B, Burack J. Enhanced perceptual functioning in autism: an update and eight principles of autistic perception. *Journal of Autism and Developmental Disorders* 2006;**36**(1):27–43.

Mundy 1994

Mundy P, Sigman M, Kasari C. Joint attention, developmental level and symptom presentation in autism. *Development and Psychopathology* 1994;**6**(3):389–401.

Nassar 2009

Nassar N, Dixon G, Bourke J, Bower C, Glasson E, De Klerk N, et al. Autism spectrum disorders in young children: effect of changes in diagnostic practices. *International Journal of Epidemiology* 2009;**38**(5):1245–54.

Ozonoff 1991a

Ozonoff S, Pennington B, Rogers SJ. Executive function deficits in high-functioning autistic individuals: relationship to theory of mind. *Journal of Child Psychology and Psychiatry* 1991;**32**(7):1081–105.

Ozonoff 1991b

Ozonoff S, Rogers SJ, Pennington B. Asperger's syndrome: evidence of an empirical distinction from high-functioning autism. *Journal of Child Psychology and Psychiatry* 1991;**32**(7):1107–22.

Parsons 1999

Parsons S, Mitchell P. What children with autism understand about thoughts and thought bubbles. *Autism* 1999;**3**(1):17–38.

Ponnet 2004

Ponnet KS, Roeyers H, Buysse A, De Clercq A, Van Der Heyden E. Advanced mind reading in adults with Asperger syndrome. *Autism* 2004;**8**(3):249–66.

Premack 1978

Premack D, Woodruff G. Does the chimpanzee have a theory of mind?. *Behavioural and Brain Sciences* 1978;**1**(4):515–26.

Ruffman 2001

Ruffman T, Garnham W, Rideout P. Social understanding in autism: eye gaze as a measure of core insights. *Journal of Child Psychology and Psychiatry* 2001;**42**(8):1083–94.

Rutter 2003

Rutter M, Bailey A, Berument SK, Lord C, Pickles A. *Social Communication Questionnaire*. Los Angeles, USA: Western Psychological Services, 2003.

Rutter 2005

Rutter M. Aetiology of autism: findings and questions. *Journal of Intellectual Disability Research* 2005;**49**(Pt 4): 231–8.

Schopler 1986

Schopler E, Reichler RJ, Renner BR. *The Childhood Autism Rating Scale (CARS): For Diagnostic Screening and Classification of Autism*. New York, USA: Irvington, 1986.

Swettenham 2000

Swettenham J. Teaching theory of mind to individuals with autism. In: Baron-Cohen S, Tager-Flusberg H, Cohen DJ editor(s). *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience*. Oxford: Oxford University Press, 2000:442–58.

Tager-Flusberg 2000

Tager-Flusberg H. Language and understanding minds: connections in autism. In: Baron-Cohen S, Tager-Flusberg H, Cohen DJ editor(s). *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience*. 2nd Edition. Oxford, UK: Oxford University Press, 2000: 124–49.

Wellman 2000

Wellman HM, Lagattuta KH. Developing understandings of mind. In: Baron-Cohen S, Tager-Flusberg H, Cohen DJ editor(s). *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience*. 2nd Edition. Oxford, UK: Oxford University Press, 2000:21–49.

Wellman 2001

Wellman HM, Cross D, Watson J. Meta-analysis of theory of mind development: the truth about false belief. *Child Development* 2001;**72**(3):655–84.

Williams 2006

Williams JG, Higgins JPT, Brayne CEG. Systematic review of prevalence studies of autism spectrum disorders. *Archives of Disease in Childhood* 2006;**91**(1):8–15.

Yeargin-Allsopp 2003

Yeargin-Allsopp M, Rice C, Karapurkar T, Doernberg N, Boyle C, Murphy C. Prevalence of autism in a US metropolitan area. *JAMA* 2003;**289**(1):49–55.

References to other published versions of this review**Fletcher-Watson 2010**

Fletcher-Watson S, McConachie H. Interventions based on the Theory of Mind cognitive model for autism spectrum disorder (ASD). *Cochrane Database of Systematic Reviews* 2010, Issue 10. [DOI: 10.1002/14651858.CD008785]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies [ordered by study ID]

Baghdadli 2013

Methods	RCT. Control condition: structured group leisure activities
Participants	High-functioning autism Aged eight to 12 years All male Intellectual level - normal range. Participants matched on verbal IQ Diagnosis by clinical best estimate, confirmed by ADOS and ADI-R Final n = 13 (93%)
Interventions	Designed to teach ToM and precursor skills Targets emotion recognition from faces, conversation, social problem solving, stress management, and ToM Delivery in group treatment sessions with two therapists. Methods include role play, video-modelling and problem solving skills. Based on manual of treatment goals including techniques, learning stages and tools Dose: 90 minutes per week, 20 weeks spread over six months
Outcomes	Primary outcome: DANVA-2 short form (emotion recognition assessment) Secondary measures: KidScreen, parent-report measure of quality of life
Notes	This study was partially funded by the Pfizer Foundation.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computerised random number generator
Allocation concealment (selection bias)	Unclear risk	List given to therapists, though allocation did take place after baseline assessment
Blinding of participants and personnel (performance bias) All outcomes	High risk	"Until the end of the study, none of the participants (children and parents) were told which group they were assigned to" (p. 436). However not possible to blind them to group activity content
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors of outcome measures were blinded

Baghdadli 2013 (Continued)

Incomplete outcome data (attrition bias) All outcomes	Low risk	Use Intention-to-treat analysis as well as assessment of study-completers
Selective reporting (reporting bias)	Low risk	No evidence of selective reporting
Other bias	Low risk	Used independent fidelity coding system to ensure treatment adherence by therapists

Begeer 2011

Methods	RCT; wait-list control
Participants	All forms of ASD Aged eight to 13 years old Both genders Intellectual level: normal range. Participants matched on chronological age, full-scale IQ, verbal IQ, and non-verbal IQ Diagnosis by clinical best estimate confirmed by SRS and ASQ Final n = 36 (90%)
Interventions	Designed to teach ToM Target ToM understanding Delivered by manualised small group training, plus monthly parental training and homework. Sessions supervised by certified therapists. Training progresses through 53 structured settings from precursors of ToM (e.g. recognising emotions and intentions), through elementary skills (e.g. deception, understand others' mental states) to complex ToM skills (e.g. second order mental state reasoning) Dose: one and a half hours weekly for 16 weeks
Outcomes	Primary outcome: The ToM test. Secondary measures: Levels of Emotional Awareness Scale - Children; self-reported empathy; parent-report social behaviours
Notes	Funding source not reported.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	"An independent researcher made the allocation schedule", p. 1000
Allocation concealment (selection bias)	Unclear risk	Insufficient information available
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible to blind personnel and participants

Begeer 2011 (Continued)

Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	No description of outcome assessors
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Four families lost to follow-up, three from control group: “re-fused to return for testing”, p. 1000
Selective reporting (reporting bias)	Low risk	No evidence of selective reporting
Other bias	Low risk	

Bolte 2002

Methods	RCT: wait-list control
Participants	HFA and AS Aged 16 to 40 years All male Intellectual level: normal range. Participants matched on non-verbal IQ Diagnosis confirmed by ADOS & ADI-R Final n = 10 (100%)
Interventions	Designed to teach precursor skills of ToM Targets affect recognition from face and eyes Delivered by “FEFA” computer program: presents six basic emotions plus neutral faces, users match faces or pictures of eyes to the correct written label Dose: two hours per week for five weeks
Outcomes	Primary outcome measures: Built-in test condition and IAPS (International Affective Picture System) ratings Secondary outcome: fMRI
Notes	Bolte 2006 - second report of same data set. This study was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG)

Risk of bias

Bias	Authors’ judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Participants pooled and numbered randomly, then allocated alternately to each group (information provided by email)
Allocation concealment (selection bias)	Low risk	Participants assigned to group after initial assessment, using procedure described above

Bolte 2002 (Continued)

Blinding of participants and personnel (performance bias) All outcomes	High risk	Participants and personnel not blind
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Outcome assessors not blind, but primary outcomes measures computer administered (p. 65)
Incomplete outcome data (attrition bias) All outcomes	Low risk	No evidence of attrition
Selective reporting (reporting bias)	High risk	IAPS data not reported in either publication; author attempted to locate these scores but no longer available fMRI loss of data in 2006 report unexplained.
Other bias	Low risk	

Fisher 2005

Methods	RCT; executive function training comparison group, also TAU control but non-random selection for this group	
Participants	ASD and AS. Aged six to 15 years old. Both genders. Intellectual level: low range. Participants matched on verbal IQ and non-verbal IQ raw scores Diagnosis based on clinical best estimate plus DSM-IV checklist Final n = 27 (100%).	
Interventions	Designed to teach ToM. Targets false belief-skills. One-to-one training by therapists, taught a strategy of thinking about beliefs as “photos in the head”, and using illustrative stories. Progress across five stages of increasing complexity existence of mental states to comprehending false belief Dose: up to a maximum of 10 x 25mins (p. 763).	
Outcomes	Primary Outcome: ToM ability, measured by false belief and other tasks, including teacher-report measure	
Notes	This study was funded by the Medical Research Council, UK.	
<i>Risk of bias</i>		
Bias	Authors' judgement	Support for judgement

Fisher 2005 (Continued)

Random sequence generation (selection bias)	High risk	Participants in control group selected by non-attendance at school during training period. Allocation to ToM and executive function intervention groups may be random, but no method detail provided (p. 759). Also “children from the same school were distributed across groups, to control for any school effects” (p. 759)
Allocation concealment (selection bias)	High risk	Not addressed in the report, but control group not randomly assigned
Blinding of participants and personnel (performance bias) All outcomes	High risk	Impossible to conceal intervention type from therapist or participant but participants probably not aware of link between intervention type and different outcome measures
Blinding of outcome assessment (detection bias) All outcomes	High risk	Unclear but seems likely that outcome tests were administered by therapists
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Five teacher questionnaires are missing (p. 759) attributed to the fact that they had too many to complete each day
Selective reporting (reporting bias)	Low risk	No evidence of selective reporting.
Other bias	Low risk	

Golan 2006

Methods	RCT; TAU control.
Participants	HFA & AS. Aged 17 to 52 years. Both genders. Intellectual level: normal range. Participants matched on chronological age, full-scale, verbal and non-verbal IQ Diagnosis confirmed by AQ. n = 41 (89%).
Interventions	Designed to teach a precursor skill of ToM. Targets complex emotion recognition. Delivered by MindReading software at home: an emotion library, trainer, and games all promote understanding of how photographs and film clips of facial emotions match on to vocabulary and emotional stories Dose: required minimum use of 10 hours over a 10-week period

Golan 2006 (Continued)

Outcomes	Primary outcome: Emotion recognition using a range of computerised tests
Notes	This study was funded by multiple contributors comprising: the National Alliance for Autism Research, the Corob Charitable Trust, the Cambridge Overseas Trust, the B'nai B'rith Leo Baeck scholarships, the Shirley Foundation, the Medical Research Council and the Three Guineas Trust

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	States random allocation but not described.
Allocation concealment (selection bias)	High risk	Not described directly but report implies randomisation occurred before initial assessment
Blinding of participants and personnel (performance bias) All outcomes	High risk	Participants not blind but unaware of intervention goals of the study
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Three assessors were blind to group but one (the lead author) was not. Outcome measures were computer-presented and therefore resilient to bias
Incomplete outcome data (attrition bias) All outcomes	High risk	Five participants in the intervention group dropped out due to not finding time to complete the required amount of work (p. 597). Study may over-estimate value of intervention
Selective reporting (reporting bias)	Low risk	One participant's data not reported for one of many outcome measures (p. 602, Table 2)
Other bias	Low risk	

Golan 2010

Methods	RCT; TAU control.
Participants	ASD. Aged four to eight years. Both genders. Intellectual level: normal range. Participants matched for chronological age and verbal IQ Diagnosis confirmed by ADI-R and CAST. Final n = 38 (97%).

Golan 2010 (Continued)

Interventions	Designed to teach precursor skills of ToM Targets emotion comprehension Delivered by <i>Transporters</i> cartoon on DVD watched at home: cartoon episodes depict vehicles with real human faces who experience emotional situations and display appropriate facial reactions Dose: recommend three episodes per day for four weeks.
Outcomes	Primary Outcomes: Emotion vocabulary and emotion recognition from complex situations
Notes	This study was funded by Culture Online and the UK Department for Culture, Media and Sport

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	States random allocation but procedure not described.
Allocation concealment (selection bias)	Unclear risk	Timing and method not described.
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible to blind participants.
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Blinding of outcome assessors not reported, but computer presentation of outcomes measures resistant to bias
Incomplete outcome data (attrition bias) All outcomes	Low risk	Only one participant dropped out (control group) but this is unexplained
Selective reporting (reporting bias)	Low risk	No evidence of this.
Other bias	Low risk	

Goods 2013

Methods	RCT: TAU control (regular school program, 30 hrs per week).
Participants	Autism diagnosis. Aged three to five years. No gender information provided. Intellectual level: low range. Participants matched on chronological age, full-scale IQ, and verbal IQ Clinical best estimate diagnosis confirmed by ADOS. Final n = 11 (73%).

Goods 2013 (Continued)

Interventions	Designed to teach a precursor skill of ToM. Targets joint engagement and social communication skills. Delivered by individual therapists sessions: children taken out of regular classroom activities. Approach also described in Kasari 2006 & Kasari 2010 . Dose: Two sessions of 30 minutes each per week, for 12 weeks intervention period	
Outcomes	Primary outcomes: Early Social Communication Scales, Structured Play Assessment, classroom observation of joint engagement, and spontaneous communicative gestures	
Notes	This study was funded by the Organisation for Autism Research and Autism Speaks	
<i>Risk of bias</i>		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Blocked randomisation using SPSS random number generator.
Allocation concealment (selection bias)	High risk	Allocation post-assessment but concealment not described. As the blocks were of fixed length the final allocation within each block would be known
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible to blind participants or personnel.
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors blinded for all outcomes.
Incomplete outcome data (attrition bias) All outcomes	High risk	Attrition four out of 15 participants. Analysis considered completers only
Selective reporting (reporting bias)	High risk	One measure (Reynell Developmental Language Scales) reported in Table 2 (no group differences at Baseline or Exit) but not discussed
Other bias	Low risk	

Hadwin 1996

Methods	RCT: three intervention groups (emotion, belief, play).
Participants	Autism. Aged four to 13 years. Both genders. Intellectual level: mostly low range, but four children with verbal mental age in normal range (distributed across groups). Participants matched on chronological age, verbal mental age, and expressive language Diagnosis confirmed by DSM-III and DSM-IV checklist (see Hadwin 1997 , p. 524). Final n = 30 (100%).
Interventions	Designed to teach ToM. Three intervention groups each targeting a specific ToM component: Understanding Emotion, Understanding Belief, and Pretend Play Delivered by one-to-one therapist led intervention: used a question and answer structure with corrective feedback. Play session additional involved modelling of pretend play. Good performance on each level required for progress to the next level of complexity Dose: One assessment session, then eight therapy sessions, followed by one post-test assessment session. Eight, consecutive daily half-hour therapeutic sessions
Outcomes	Primary outcome: teaching level reached on five-point scale. Secondary outcomes: generalisation of taught skills to novel materials. Conversation and use of mental state language in a story-book task (Hadwin 1997).
Notes	also see Hadwin 1997 . This study was funded by the Bethlem-Maudsley Research Fund and the Mental Health Foundation

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	States random allocation but procedure not described.
Allocation concealment (selection bias)	High risk	Baseline assessment conducted after allocation to treatment condition
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible to blind participants or personnel.
Blinding of outcome assessment (detection bias) All outcomes	High risk	Not explicitly stated, but seems that both pre- and post-test data were collected by unblinded therapists
Incomplete outcome data (attrition bias) All outcomes	Low risk	No evidence of attrition.

Hadwin 1996 (Continued)

Selective reporting (reporting bias)	Low risk	No evidence of selective reporting.
Other bias	Low risk	

Hopkins 2011

Methods	RCT: control group with art and design computer program.
Participants	ASD. Aged six; three to 15; one years. Both genders. Intellectual level: 24 in the normal range (HFA), 25 in a low-functioning range (LFA). Participants matched on chronological age, verbal IQ and non-verbal IQ Diagnosis by clinical best estimate confirmed with CARS. Final n = 49 (96%).
Interventions	Teaches precursor skills of ToM. Targets emotion recognition, eye contact and facial identity recognition Delivered by FaceSay: social skills training via a computer program: three games to: match gaze direction with an object; select parts of faces which fit with a whole face; and match facial expressions with emotions Dose: 12 x 30 minute sessions over 6 weeks.
Outcomes	Primary outcome: Emotion recognition from photos and drawings of faces and Benton Face Recognition Test Secondary outcomes: Social Skills Rating System, Social skills observation
Notes	<i>Authors were emailed for further information, no response.</i> This study was partially funded by a grant from Civitan International and with the cooperation of Symbionica LLC

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	No information provided.
Allocation concealment (selection bias)	Unclear risk	No information provided.
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	Parents and participants technically blind but could have guessed from computer program content
Blinding of outcome assessment (detection bias) All outcomes	Low risk	SSRS: observed by blinded raters. Other measures resistant to bias.

Hopkins 2011 (Continued)

Incomplete outcome data (attrition bias) All outcomes	Low risk	Two children excluded due to low attendance, and reasons given
Selective reporting (reporting bias)	Low risk	No evidence of selective reporting.
Other bias	Low risk	

Ingersoll 2012

Methods	RCT; TAU control.	
Participants	Core autism. Aged 27 to 47 months old. Both genders. Intellectual level: low range. Participants matched on chronological age, non-verbal mental age, and language age Diagnosis based on ADOS and clinical best estimate. Final n = 27 (93%).	
Interventions	Designed to teach precursor skills of ToM. Targets imitation skills. Therapist-delivered reciprocal imitation training: naturalistic play-based imitation intervention using pairs of identical play materials to model social imitation skills. Actions are modelled and paired with a verbal marker with physical prompts used where children do not respond Dose: three hours per week for 10 weeks plus.	
Outcomes	Primary outcomes: Early Social Communication Scales, initiation of joint attention. Social Emotional Scale of the Bayley Scales of Infant Development 3rd Edition, parent-report Secondary (mediator) measures: Elicited and spontaneous imitation skills (motor imitation, unstructured imitation assessment)	
Notes	See also Ingersoll 2010. Funding source not known.	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Randomised by a coin toss.
Allocation concealment (selection bias)	Low risk	Pairing on expressive language age before assignment by coin toss

Ingersoll 2012 (Continued)

Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible to blind participants and therapists.
Blinding of outcome assessment (detection bias) All outcomes	Low risk	In Ingersoll 2010, p. 1155 “Assessments were scored by trained research assistants blind to group” Confirmed in personal communication: although examiners were not blind, coders were
Incomplete outcome data (attrition bias) All outcomes	Low risk	Two children withdrew, and reasons given. Analysis of completers only
Selective reporting (reporting bias)	Low risk	No evidence of selective reporting.
Other bias	Unclear risk	No reason given for adding additional subjects to Ingersoll 2010, to randomisation may not be secure

Kaale 2012

Methods	RCT: TAU control.
Participants	Autism and ASD. Aged 29 to 60 months. Both genders. Intellectual level: low range. Participants matched on chronological age, mental age, expressive and receptive language ages, and developmental level Diagnosis confirmed by ADOS (or ADI for 49 participants). Final n = 61 (100%).
Interventions	Designed to teach precursor skills of ToM. Targets joint attention and joint engagement. Delivered as manualised Joint attention intervention, by pre-school teachers: table top training sessions provide multiple opportunities for a child to initiate a targeted skill. Skills are encouraged using verbal and physical prompts, presenting opportunities and positive feedback Dose: two daily sessions, five days per week for eight weeks
Outcomes	Primary outcome: ESCS. Secondary outcome: Teacher-child play and mother-child, both rated for joint attention and joint engagement
Notes	This study was funded by the South-Eastern Norway Regional Health Authority and the Centre for Child and Adolescent Mental Health (Eastern and Southern Norway)

Risk of bias

Bias	Authors' judgement	Support for judgement
------	--------------------	-----------------------

Kaale 2012 (Continued)

Random sequence generation (selection bias)	Low risk	Concealed randomisation list generated before the study began
Allocation concealment (selection bias)	Low risk	Allocation after baseline assessment.
Blinding of participants and personnel (performance bias) All outcomes	High risk	Parents and pre-school staff were blind to allocation at baseline (p. 98), but therapists were not blind (p. 101). Participants likely to be unaware of intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Blinded coders and outcome assessors (p. 98).
Incomplete outcome data (attrition bias) All outcomes	Low risk	No evidence of attrition.
Selective reporting (reporting bias)	Low risk	All pre-specified outcomes were reported.
Other bias	Low risk	

Kasari 2006

Methods	RCT; two control groups: TAU and symbolic play.
Participants	ASD and core autism. Aged 36 to 60 months. Both genders. Intellectual level: low range. Participants matched on chronological age, mental age, developmental level and expressive and receptive language ages Diagnosis confirmed by ADOS and ADI. Final n = 58 (89%).
Interventions	Designed to teach precursor skills of ToM. Targets joint attention. Delivered as one to one therapist led intervention: table top training and then floor sessions provide multiple opportunities for a child to initiate a targeted skill. Skills are encouraged using verbal and physical prompts, presenting opportunities and positive feedback. Structure is withdrawn as the training progresses Dose: 30 minutes daily for five to six weeks.
Outcomes	Primary outcomes: ESCS. Secondary outcome: parent-child play rating, coded for joint engagement and joint attention
Notes	See also: Gulsrud 2007 : sub-analysis of n = 35 with age range 33 to 54 months. Arora 2008 : PhD thesis reporting on links between joint attention and perseveration for n = 35

Kasari 2006 (Continued)

	<p>Kasari 2008: specific report on language outcome for the full sample, Reynell scales of language development Kasari 2012b: longitudinal follow-up, reporting on expressive vocabulary and Differential Abilities Scale This study was funded by the National Institute for Health and the CPEA network</p>	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"used generated numbers from computer program" (information provided by author)
Allocation concealment (selection bias)	High risk	"Families who consented were randomised and began treatment right after assessments (by outside independent assessors" and "allocation to group was done as kids entered the Early Intervention Program and parent consented - next allocation of group number was revealed after eligibility and consenting" (information provided by author) . However the published report states that some children were excluded because they did not meet inclusion criteria (p. 612) indicating that assessments were completed but not scored before allocation
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible to blind parents, children or therapists, though children likely to have been unaware
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Outcome assessors and coders both blind to group (information provided by author)
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Seven participants did not meet criteria, refused final assessment or left the program unexpectedly. Distributed across groups (Joint attention intervention = two; symbolic play intervention = one; control group = four)
Selective reporting (reporting bias)	Unclear risk	No evidence of selective reporting.
Other bias	Low risk	

Kasari 2010

Methods	RCT: wait-list control.
Participants	Core autism. Aged 21 to 36 months. Both genders. Intellectual level: low range. Participants matched on chronological age, mental age, and developmental level Diagnosis by clinical best estimate and ADI. Final n = 35 (83%).
Interventions	Designed to teach precursor skills of ToM. Targets joint attention and joint engagement. Delivered as parent-mediated training adapted from Kasari 2006 . Core principles form 10 modules, delivered by parents trained by therapists. Uses aspects of applied behaviour analysis and facilitative and responsive interaction to encourage target behaviours in the child Dose: three modules per week for eight weeks.
Outcomes	Primary outcome: Videotaped parent-child interaction, coded for engagement, type of play, and frequency of joint attention Secondary outcomes: caregiver quality of involvement, adherence to treatment, and service utilisation measure
Notes	This study was funded by the National Institute for Mental Health

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random numbers list.
Allocation concealment (selection bias)	Low risk	Children randomised after meeting study criteria.
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible to blind parents and participating children.
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Examiners and coders both blind to group status.
Incomplete outcome data (attrition bias) All outcomes	Low risk	Two children did not meet criteria, two parents declined to participate. Three further children from the control group described as "did not receive allocated intervention". Though not explicitly stated it appeared from Table 4 that analyses were based on intention-to-treat rather than treatment completers alone
Selective reporting (reporting bias)	Low risk	No evidence of selective reporting (pp. 1052-1053).

Other bias	Low risk
------------	----------

Kim 2009

Methods	RCT with cross-over; toy play session control.
Participants	Core autism. Aged 39 to 71 months. Final sample all male. Intellectual level: normal range. Participants not matched but evaluated for developmental level Diagnosis confirmed by Korean CARS and four received ADOS as well Final n = 10 (66%).
Interventions	Designed to teach precursor skills of ToM. Targets joint attention. Delivered as improvised music sessions based on a semi-flexible treatment manual. Sessions included free play time and directed activities with therapist modelling turn-taking and other activities Dose: 12 weeks, 30 minutes per week.
Outcomes	Primary outcomes: PDD-BI social approach subscale, and ESCS. Secondary outcomes: observational coding of emotional and motivation responsiveness, and responsiveness to joint attention / joint attention initiation
Notes	See also Kim 2008 for details of PDD-BI and ESCS outcome measure. <i>Email sent to request more information but none received.</i> This study was funded by Aalborg University, Denmark.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	"participants were randomly assigned to two groups" - but no details given
Allocation concealment (selection bias)	Unclear risk	No details given.
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible to blind participants and personnel.
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Coders for PDD-BI and ESCS both blind. PDD-BI administered to both blind teachers and unblinded parents, so in the latter case, susceptible to bias even if the interviewer is blind 30% of observational behaviour measures

Kim 2009 (Continued)

		second-coded by blind coder for reliability
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Five children dropped out due to ill health or long distances to travel to receive intervention
Selective reporting (reporting bias)	Low risk	No evidence of selective reporting.
Other bias	Low risk	Differences in ratings provided by mothers and professionals. Professional ratings support efficacy of intervention and these were also blind. Mothers over-estimate pre-intervention abilities and thus under-estimate efficacy of the intervention, but they were not blind (in Kim, 2008, p. 1763)

Landa 2011

Methods	RCT: contact (non-interpersonal synchrony) control.	
Participants	Autism and ASD. Aged 21 to 33 months. Both genders. Intellectual level: low range. Participants matched on chronological age, receptive language and visual reception standardised scores Diagnosis confirmed by ADOS. Final n = 48 (96%).	
Interventions	Designed to teach precursor skills of ToM. Targets socially synchronous behaviour. Delivered by trained teacher using the Assessment Evaluation Program System (AEPS) for infants and children. Instructional strategies ranged from discrete trial teaching to pivotal response training and routine-based interactions. Intervention incorporated low-tech communication systems. Highly motivating tasks and materials designed to elicit frequent child-initiated communication Dose: two and a half hours per day, four days per week for six months in classroom Additional one and a half hours per month parent training and 38 hours parent education	
Outcomes	Primary Outcome: Socially engaged imitation in a structured imitation task Secondary outcomes: Initiations of joint attention & shared positive affect - measured using the CSBS; Mullen Scales of Early Learning expressive language and visual reception subscales	
Notes	This study was funded by the National Institute of Mental Health and HRSA	
<i>Risk of bias</i>		
Bias	Authors' judgement	Support for judgement

Landa 2011 (Continued)

Random sequence generation (selection bias)	Unclear risk	No information provided.
Allocation concealment (selection bias)	Unclear risk	Participants placed into matched pairs prior to randomisation which could introduce bias
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Parent, children and teachers all blind to group: not possible to blind intervention trainers
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessment at baseline and follow-up by a independent, blind clinician (p. 16)
Incomplete outcome data (attrition bias) All outcomes	Low risk	One family withdrew from the study, and one was not analysed due to missing baseline data
Selective reporting (reporting bias)	Low risk	No evidence of selective reporting.
Other bias	Low risk	

Ryan 2010

Methods	RCT: wait-list control.
Participants	Autism and ASD. Aged six to 14 years. Both genders. Intellectual level: normal range. Participants matched on verbal IQ and non-verbal IQ Diagnosis confirmed by ADOS and DISCO. Final n = 30 (100%).
Interventions	Designed to teach precursor skills of ToM. Targets emotion recognition. Delivered as therapist-led training using photos of six core facial emotions. Therapists highlighted component parts of emotional expressions (e.g. raised eye-brows) and matched faces with verbal labels. Participants engaged in matching games, role-play, tracing, and drawing of faces to support direct instruction Dose: one hour per week over four weeks.
Outcomes	Primary outcome: Facial emotion recognition test.
Notes	This study was funded by the COPE Foundation and supported by the Catherine T MacArthur Foundation Research Network on Early Experience and Brain Development
<i>Risk of bias</i>	

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"We assigned sequential numbers to each child (ordered alphabetically) to be included in the study and then used the random number table provided in David Clark-Carter's book "Doing Quantitative Psychological Research - From Design to Report" (1997) to form two groups" (information provided by author)
Allocation concealment (selection bias)	High risk	Assessments conducted before allocation, but using open random numbers table could have introduced bias
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible to blind personnel and participants.
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Independent outcome assessor was "blind to the initial scores achieved by each of the children, to the membership of the treatment groups and to our hypotheses" (information provided by author)
Incomplete outcome data (attrition bias) All outcomes	Low risk	No evidence of attrition at four weeks, immediately post intervention. However five children lost to follow-up at three months, after all had received the training, without reason given
Selective reporting (reporting bias)	Low risk	No evidence of selective reporting, but only one outcome measure was used
Other bias	Low risk	

Schertz 2013

Methods	RCT: TAU control.
Participants	Autism spectrum disorder and autism. Aged under 30 months (mean age in each group: 24.6 months and 27.5 months) Gender information not reported. Intellectual level: normal range. Participants matched on chronological age, receptive and expressive language Diagnosis confirmed by ADOS and M-CHAT (to establish high risk) Final n = 23 (100%, but see risk of bias below).
Interventions	Designed to teach precursor skill of ToM. Targets joint attention and sub-skills (Focus on Faces, Turn-Taking) Delivered as parent-training delivered by intervention co-ordinators at home. Sessions included guidance and reflection based on filmed segments of parent-child play, and a manual for parents

Schertz 2013 (Continued)

	Dose: Parent training sessions were weekly at home for at least 15 sessions. Parents agree to spend 30 minutes per day in face-to-face interaction with their child. Total intervention period ranged from 4 to 12 months (mean 7 months)	
Outcomes	Primary outcomes: Precursors of Joint Attention Measure (PJAM), based on coding of parent-child interaction Secondary measures: Vineland Adaptive Behaviour Scales and Mullen Scales of Early Learning	
Notes	This study was funded by Autism Speaks.	
<i>Risk of bias</i>		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Once entry criteria were met, children were paired and then assigned to intervention or control. Randomisation was by coin toss (confirmed by author email)
Allocation concealment (selection bias)	High risk	No information given, but coin toss open to abuse if performed by intervention co-ordinators
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible to blind parents and intervention co-ordinators
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Intervention co-ordinators and parents carried out assessments but these were coded by raters blind to group (but not time point)
Incomplete outcome data (attrition bias) All outcomes	Low risk	There is unreported attrition of one participant (confirmed by author)
Selective reporting (reporting bias)	Unclear risk	No evidence of selective reporting.
Other bias	High risk	The treatment period was of variable length. Treatment was terminated (and outcomes measures taken) once three or more instances of initiation of joint attention were observed across multiple sessions. Within each pair, inter-assessment period was identical, but this variable system means that 100% of participants made gains in responses to treatment. i.e. the treatment effect was weighted by this study design feature

Solomon 2004

Methods	RCT, wait-list control.
Participants	HFA, AS and PDD-NOS. Aged eight to 12 years. All male. Intellectual level: normal range, Participants matched on chronological age and full-scale IQ Diagnosis confirmed by ADOS and ADI. Final n = 18 (100%).
Interventions	Designed to teach ToM. Targets social adjustment, emotion recognition, ToM, and executive functions Delivered as group social skills training sessions with parent psychoeducational sessions. Children's group sessions followed a strict structure, including welcome, lesson time, games and free social time. The curriculum followed 10 modules focusing on emotional understand and empathy, conversational skills, and lessons about friendship Dose: one and a half hours per week for 20 weeks.
Outcomes	Primary outcomes: Diagnostic Analysis of Non-Verbal Accuracy - facial expression recognition subscale; Happe's Strange Stories; The Faux-Pas Recognition Test; Test of Problem Solving (TOPS) Secondary outcomes: Depression Inventories for children and parents; problem behaviour logs
Notes	<i>Authors were emailed for further information but no response received.</i> This study was funded by the MIND Institute, UC Davis, California, USA

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Participants were matched into pairs and then "Nine boys were randomly selected... to serve as the first intervention group subjects" - but method unclear (p. 654)
Allocation concealment (selection bias)	Unclear risk	Participants assessed and matched before random allocation, but pairing process could have introduced bias
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible to blind participants and personnel.
Blinding of outcome assessment (detection bias) All outcomes	High risk	Unlikely to have been blinded as assessments carried out by first author
Incomplete outcome data (attrition bias) All outcomes	Low risk	No evidence of attrition.

Solomon 2004 (Continued)

Selective reporting (reporting bias)	Low risk	No evidence of selective reporting.
Other bias	Low risk	

Williams 2012

Methods	RCT; control group Thomas the Tank Engine DVD.
Participants	Autism spectrum disorder, including autism. Aged four to seven years. Intellectual level: low to normal range, but cognitive level above 30 months minimum. Participants matched on chronological age, full-scale IQ, verbal IQ and non-verbal IQ Diagnosis based on clinical best estimate confirmed by ADOS. n = 55 (92%).
Interventions	Designed to teach a precursor skill of ToM. Targets emotion recognition. Delivered by <i>Transporters</i> DVD: cartoon episodes depict vehicles with real human faces who experience emotional situations and display appropriate facial reactions Dose: 15 minutes per day for four weeks (mean 11.76 hours total)
Outcomes	Primary outcomes: NEPSY-II affect recognition test using pictures, and the Pictures of Facial Affect test (using Ekman faces: identify emotions from photos of faces) Secondary measures: NEPSY-II ToM tasks; Mindreading tasks; Vineland Adaptive Behaviour Scales (socialisation subscale)
Notes	This study was funded by the Financial Markets Foundation for Children, Australia

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Statistician randomised the allocation.
Allocation concealment (selection bias)	Low risk	Each DVD packed in an unmarked numbered envelope - adequate concealment
Blinding of participants and personnel (performance bias) All outcomes	High risk	Participants not blind to DVD content.
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessor blind to treatment group.

Williams 2012 (Continued)

Incomplete outcome data (attrition bias) All outcomes	High risk	Attrition of five participants after one month (unbalanced between groups) and only completers analysed. Missing data on emotion identification where children (n = eight) could not complete the task: these may relate to outcome. Effect sizes not given. Attrition of nine more participants at three-month follow-up (unbalanced between groups) so follow-up data used last observation carried forward
Selective reporting (reporting bias)	Low risk	No evidence of selective reporting.
Other bias	Low risk	

Wong 2010

Methods	RCT cross-over: wait-list control.	
Participants	ASD. Aged 17 to 36 months. Both genders. Intellectual level: low range. Participants matched on chronological age and mental age Diagnosis confirmed by ADI-R, ADOS and CARS. Final n = 17 (100%).	
Interventions	Designed to teach precursor skills of ToM. Targets eye contact, gesture and vocalisation. Delivered as direct one to one therapy: modelling gestures accompanied by verbal labels, eliciting requests using favourite toys Dose: 10 x 30 minute sessions over two weeks.	
Outcomes	Primary outcome: ADOS (reported item by item). Secondary outcomes: Ritvo-Freeman Real Life Rating Scale (RFRLRS) - parent rated; Symbolic Play Test; Parenting Stress Index, short form	
Notes	<i>Authors emailed for more detail but no reply received.</i> Funding source unknown.	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	"17 children consecutively diagnosed were randomized into the Intervention and Control groups" - assignment may have been alternate rather than random? (p. 679)

Wong 2010 (Continued)

Allocation concealment (selection bias)	High risk	Baseline assessment carried out after randomisation to group
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible to blind participants or personnel.
Blinding of outcome assessment (detection bias) All outcomes	Low risk	ADOS and SPT both administered and scored by blind assessors. Some outcomes are parent-reported measures and therefore unblinded. (pp. 680-681)
Incomplete outcome data (attrition bias) All outcomes	Low risk	No evidence of attrition.
Selective reporting (reporting bias)	Low risk	No evidence of selective reporting.
Other bias	Unclear risk	Wait-list control group combined with Intervention group to give larger sample size for analysis

Wong 2013

Methods	Cluster-RCT: repeated measures, nested within children, nested within classrooms. Control conditions: symbolic play intervention, wait-list control
Participants	Autism. Aged three to six years. Both genders. Intellectual level: low range. Participants matched on chronological age, mental age, and expressive and receptive language ages Diagnosis based on clinical best estimate and confirmed by CARS Final n = 33 (97%).
Interventions	Designed to teach precursor skills of ToM. Targets joint attention and interpersonal engagement behaviours Delivered as teacher training. Teacher then implemented the approach in the classroom according to their own preference (e.g. one-to-one, small group, whole class) Dose: teacher training sessions were one hour per week, for four weeks
Outcomes	Primary outcome: Early Social Communication Scales, and Structured P Lay Assessment. Also direct classroom observation of child and teacher behaviours (joint engagement, joint attention) Secondary measures: teacher acceptability of intervention.
Notes	Funding source unknown.
<i>Risk of bias</i>	

Wong 2013 (Continued)

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Teacher assignment to one of three groups by random numbers list
Allocation concealment (selection bias)	Low risk	Allocation occurred before initial assessment, but allocation list concealed from baseline assessors (confirm by author email)
Blinding of participants and personnel (performance bias) All outcomes	High risk	Teachers not blinded, as recipients of training.
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Research observers for all measures were blinded (confirmed by author email)
Incomplete outcome data (attrition bias) All outcomes	Low risk	One teacher (and therefore one child) dropped out; reason given
Selective reporting (reporting bias)	Low risk	No evidence of selective reporting.
Other bias	Low risk	

Young 2012

Methods	RCT; control group Thomas the Tank Engine DVD.
Participants	All forms of ASD. Aged four to eight years. Intellectual level: normal range. Participants matched on verbal and non-verbal IQ Diagnosis based on clinical best estimate confirmed by SCQ. Final n = 25 (100%).
Interventions	Designed to teach precursor skills of ToM. Targets emotion recognition. Delivered by <i>Transporters</i> DVD: cartoon episodes depict vehicles with real human faces who experience emotional situations and display appropriate facial reactions Dose: five to 10 minute episodes, children watch three per day for three weeks
Outcomes	Primary outcome: Social Communication Questionnaire, parent-report Secondary outcomes: NEPSY-II affect recognition subscale using pictures; The Faces Task - emotion recognition from photos of faces
Notes	Funding source unknown.
<i>Risk of bias</i>	

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	"Random allocation was done by simply alternating between the 2 intervention groups for each participant as they responded to the letters which were sent out . . . we tossed a coin to see which group went first and it was Thomas and then the next person who responded was Transporters and so on" (information from the author)
Allocation concealment (selection bias)	High risk	Published report states that assignment was after baseline assessment (p. 987) but author's personal communication states that "random allocation occurred before baseline assessment- when the participant registered interest in the study they were allocated a condition"
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Parents and children aware of DVD content but "they were naive to the purpose of the study" (information from the author)
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	"assessor was not blind to treatment but the measures didn't require any subjective interpretation. Parents completed questionnaires and they were naive to the purpose of the study" (information from the author)
Incomplete outcome data (attrition bias) All outcomes	Low risk	No evidence of attrition.
Selective reporting (reporting bias)	Low risk	All pre-specified outcomes are reported. So too are four outcome measures not pre-specified, nor described in the methods
Other bias	Low risk	

ADI-R: Autism Diagnostic Interview - Revised
 ADOS: Autism Diagnostic Observation Schedule
 AS: Asperger's syndrome
 ASD: autism spectrum disorder
 ASQ: autism spectrum quotient
 CARS: Childhood Autism Rating Scale
 DANVA-2: Diagnostic Analysis of Non-Verbal Accuracy
 DISCO: Diagnostic Interview for Social and Communication Disorders
 DSM: Diagnostic and Statistical Manual of Mental Disorders
 ESCS: Early Social Communication Scales

fMRI: functional magnetic resonance imaging
HFA: high-functioning autism
ICD-10: International Classification of Diseases (10th revision)
LFA: low-functioning autism
M-CHAT: Modified Checklist for Autism in Toddlers
NEPSY-II: developmental NEuroPSYchological assessment, 2nd Edition
PDD-BI: Pervasive Developmental Disorder Behavior Inventory
PDD-NOS: Pervasive Developmental Disorder - Not Otherwise Specified
RCT: randomised controlled trial
SCQ: Social Communication Questionnaire
SPT: Symbolic Play Test
SRS: Social Responsiveness Scale
ToM: Theory of Mind

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Adams 2012	Not ToM-linked
Aldred 2004	Broad-base intervention, not ToM-specific
Beaumont 2008	Broad-base intervention, not ToM-specific
Cardon 2011	Used a multiple baseline case-control design
Carter 2011	Not ToM-linked
Casenhiser 2011	Broad-base intervention, not ToM-specific
Castorina 2011	Not ToM-linked
Charman 2007	No data reported
Corbett 2011	Not ToM-linked
Dawson 2010	Not ToM-linked
DeRosier 2011	Not ToM-linked
Drew 2002	Broad-base intervention, not ToM-specific
Estes 2011	No data reported
Field 2001	Experimental pilot study: not a full intervention trial
Frankel 2010a	Not ToM-linked

(Continued)

Frankel 2010b	No data reported
Gantman 2012	Not ToM-linked
Gattino 2011	Broad-base intervention, not ToM-specific
Geretsegger 2012	No data reported
Girolametto 1988	Not an ASD sample
Green 2010	Broad-base intervention, not ToM-specific
Hanley-Hochdorfer 2010	Not an RCT/qRCT
Heimann 2006	Experimental pilot study: not a full intervention trial
Hillier 2012	Not an RCT/qRCT
Howlin 1997	No data reported
Howlin 2007	Not ToM-linked
Jocelyn 1998	Not ToM-linked
Jones 2010	Not an RCT/qRCT
Kasari 2012a	Not ToM-linked
Kern 2011	Not an RCT/qRCT
Koenig 2010	Not ToM-linked
Lang 2010	No data reported
Laugeson 2012	Not an RCT/qRCT
Lawton 2012	Broad-base intervention, not ToM-specific, teaches play skills as well as joint attention
LeGoff 2004	Not an RCT/qRCT
Lerner 2012	Broad-base intervention, not ToM-specific, teaches a wide range of complex social skills
Lopata 2010	Broad-base intervention, not ToM-specific, teaches a wide range of complex social skills
Mahoney 2005	Not an RCT/qRCT
McConachie 2004	No data reported

(Continued)

McGregor 1998	Not an RCT/qRCT
Minshew 2010	No data reported
Morgan 2010	No data reported
Nefdt 2010	Not ToM-linked
Oosterling 2010	Not an RCT/qRCT
Ozonoff 1995	Not an RCT/qRCT
Pajareya 2011	Broad-base intervention, not ToM-specific
Quirnbach 2009	Not ToM-linked
Roberts 2011	Not ToM-linked
Rodgers 2012	Not ToM-linked
Rogers 2010	Not ToM-linked
Schertz 2007	Not an RCT/qRCT
Schmidt 2012	Not an RCT/qRCT
Silver 2001	ASD diagnosis unconfirmed
Smith 2000	Not ToM-linked
Smith 2004	Not an RCT/qRCT
Stichter 2001	Not an RCT/qRCT
Strain 2011	Not ToM-linked
Swettenham 1996	Not an RCT/qRCT
Tanaka 2010	Not ToM-linked
Temple 2007	No data reported
Thomeer 2012	Broad-base intervention, not ToM-specific, teaches a wide range of complex social skills
Turner-Brown 2008	Not an RCT/qRCT
Wellman 2002	Not an RCT/qRCT

(Continued)

Whalen 2003	Not an RCT/qRCT
White 2010	Not ToM-linked
Wood 2009	Not ToM-linked
Yoder 2006	Broad-base intervention, not ToM-specific
Yoder 2006b	Not ToM-linked

ASD: autism spectrum disorder

RCT: randomised controlled trial

qRCT: quasi-randomised controlled trial

ToM: Theory of Mind

Characteristics of studies awaiting assessment *[ordered by study ID]*

Baghdadli 2010

Methods	RCT
Participants	Fourteen boys aged between eight to 12 years old
Interventions	Comparison of the effect of two prosocial interventions (social skill entertainment versus general educative intervention)
Outcomes	
Notes	Awaiting translation

RCT: randomised controlled trial

DATA AND ANALYSES

Comparison 1. Treatment effects in meta-analysis

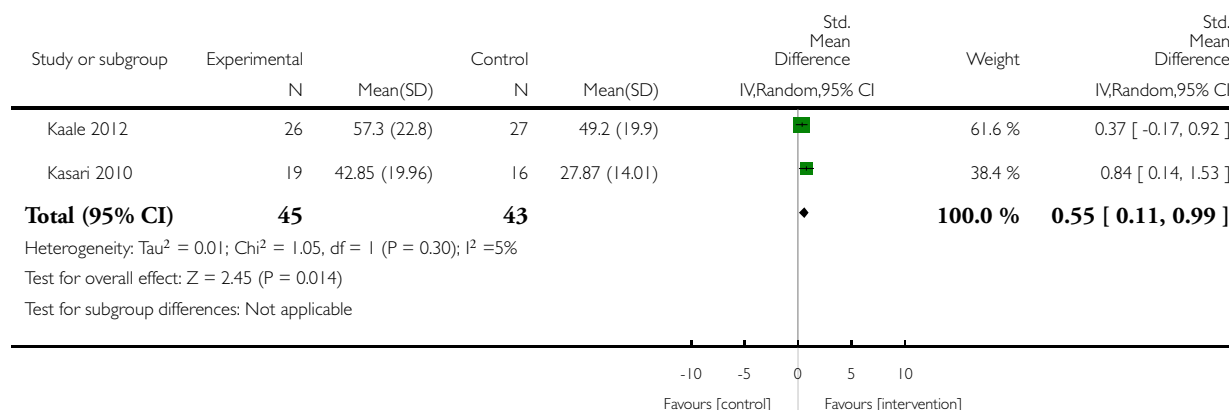
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Joint engagement in mother-child interaction	2	88	Std. Mean Difference (IV, Random, 95% CI)	0.55 [0.11, 0.99]
2 Emotion recognition from face photographs, TAU control	4	105	Std. Mean Difference (IV, Random, 95% CI)	0.75 [0.22, 1.29]
3 Joint attention initiations in standardised assessment	3	92	Std. Mean Difference (IV, Random, 95% CI)	0.23 [-0.48, 0.94]

Analysis 1.1. Comparison 1 Treatment effects in meta-analysis, Outcome 1 Joint engagement in mother-child interaction.

Review: Interventions based on the Theory of Mind cognitive model for autism spectrum disorder (ASD)

Comparison: 1 Treatment effects in meta-analysis

Outcome: 1 Joint engagement in mother-child interaction

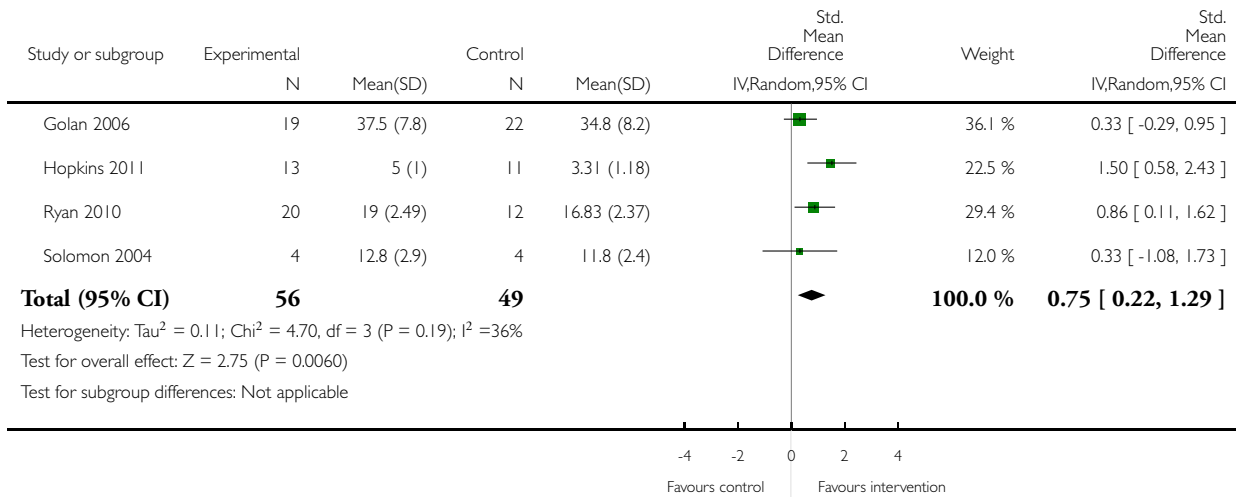


Analysis 1.2. Comparison 1 Treatment effects in meta-analysis, Outcome 2 Emotion recognition from face photographs, TAU control.

Review: Interventions based on the Theory of Mind cognitive model for autism spectrum disorder (ASD)

Comparison: 1 Treatment effects in meta-analysis

Outcome: 2 Emotion recognition from face photographs, TAU control

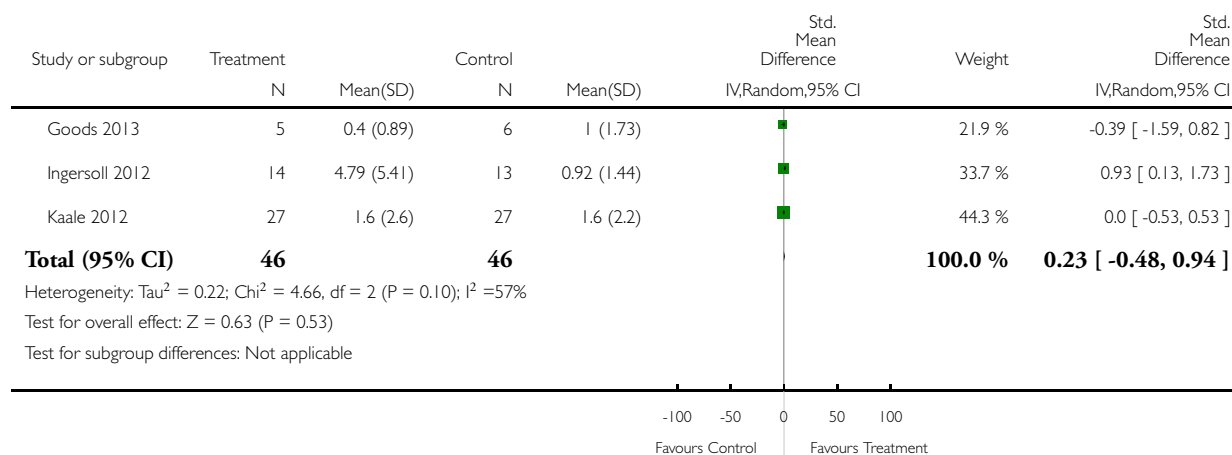


Analysis I.3. Comparison 1 Treatment effects in meta-analysis, Outcome 3 Joint attention initiations in standardised assessment.

Review: Interventions based on the Theory of Mind cognitive model for autism spectrum disorder (ASD)

Comparison: 1 Treatment effects in meta-analysis

Outcome: 3 Joint attention initiations in standardised assessment



ADDITIONAL TABLES

Table 1. Additional methods

Review section	Item	Methods
Unit of Analysis	Cluster-randomised trials	Authors will use a summary measure from each cluster and conduct the analysis at the level of allocation (that is sample size = number of clusters). However, if there are very few clusters this would significantly reduce the power of the trial, in which case the authors will attempt to extract a direct estimate of the risk ratio using an analysis that accounts for the cluster design, such as a multilevel model, a variance components analysis or generalised estimating equations (GEEs). Statistical advice will be sought to determine which method is appropriate for the particular trials to be included
Subgroup Analysis	Identification of dimensions for subgroup analysis	In future updates the following clinically-relevant differences may be the focus of subgroup analyses: 1. intervention delivery type (e.g. therapist, parent-mediated, school-based) and length

Table 1. Additional methods (Continued)

		<p>2. intervention target skill (e.g. ToM as a whole, joint attention, emotion recognition, false-belief understanding)</p> <p>3. participant age (e.g. pre-school, young children, adolescents, adults), IQ (low versus normal or high), specific diagnosis, and verbal ability</p>
Sensitivity Analysis	Identification of variables for sensitivity analysis	In future updates the impact of factors such as high rates of loss to follow-up or inadequate blinding on outcomes will be explored
Dealing with Missing Data	Procedures for imputation in the event of issues with missing data	Should unacceptable levels and/or non-random missing data be found in future studies for inclusion in the review, the authors will attempt to impute missing values. Imputation may use individual data (where available from the original report authors) OR group-level summary statistics (which are normally included in published reports). Mean imputation will be used where variables are normally distributed, and the median will be used for non-normal distributions. In either case the review will report how the imputed values appear to change the outcome of the study/meta-analysis and use this variability to inform the strength of our conclusions

Table 2. Outcome measures used

Outcome	Category	Measure	Study
Primary	Communication (standardised measure)	Semi-structured conversation task: telling a story from a picture book	Hadwin 1996
		Social Communication Questionnaire (SCQ)	Young 2012
		Autism Diagnostic Observation Schedule (ADOS) : Language and Communication	Wong 2010
	Social Function (standardised measure)	Joint Attention: Early Social Communication Scale (ESCS)	Kaale 2012 Kasari 2006 Ingersoll 2012 Goods 2013 Kim 2009 Wong 2013

Table 2. Outcome measures used (Continued)

		Communication and Symbolic Behaviour Scales developmental profile	Landa 2011
		Autism Diagnostic Observation Schedule (ADOS) : Reciprocal Social Interaction	Wong 2010
		Social Communication Questionnaire (SCQ)	Young 2012
		Social Emotional Scale (SES), Bayley Scales of Infant Development, 3rd Edition	Ingersoll 2012
		Vineland Adaptive Behaviour Scales (VABS), Socialisation subscale	Williams 2012
	Flexibility & imagination (standardised measure)	<i>none</i>	
Secondary	Intervention specific: ToM	ToM test: standardised interview for Tom understanding Levels of emotional awareness scale for children (LEAS-C): performance assessment	Begeer 2011
		False-belief tasks (unexpected transfer and deceptive box): behavioural ToM task Penny Hiding Deception Task: behavioural ToM task Seeing Leads to Knowing Task: behavioural ToM task Knowing/Guessing Task: behavioural ToM task	Fisher 2005
		Level of training reached (ToM skills, pretend play skills, emotion understanding) Generalisations to non-taught tasks (ToM skills, pretend play skills, emotion understanding) Generalisation across skill sets and intervention groups (e.g. effects of ToM intervention on pretend play skills and so on)	Hadwin 1996
		ToM: Strange Stories and Faux Pas Recognition Test	Solomon 2004

Table 2. Outcome measures used (Continued)

		NEPSY-II ToM task	Williams 2012
	Intervention specific: emotion recognition	FEFA test module: computerised facial emotion recognition test International Affective Picture System (IAPS) facial emotion recognition test fMRI evidence of change in neural response to emotional stimuli	Bolte 2002
		Diagnostic Analysis of Nonverbal Accuracy2 (DANVA2) Faces test: emotion recognition	Solomon 2004 Baghdadli 2013
		Reading-the-Mind-in-the-Eyes task: interpreting mental states from images of eyes	Fisher 2005 Golan 2006
		Cambridge MindReading face-voice battery: computerised emotion recognition test (close generalisation) Emotion recognition from novel film clips (holistic distant generalisation)	Golan 2006 Williams 2012
		Matching familiar emotional situations to familiar facial expressions (close generalisation) Matching novel emotional situations to novel facial expressions but familiar characters (unfamiliar close generalisation) Matching novel emotion situations to novel facial expressions on novel faces (distant generalisation)	Golan 2010
		Emotion Recognition Test: photographs of faces Emotion Vocabulary Comprehension Test	Ryan 2010
		Recognition of emotional expressions from photographs Recognition of emotional expressions from line drawings Benton Facial Recognition Test (short form)	Hopkins 2011
		NEPSY: Affection Recognition subtest (recognising emotions from photos of faces)	Williams 2012 Young 2012

Table 2. Outcome measures used (Continued)

		The Faces Task: recognising emotions from photos of faces	
	Intervention specific: imitation	Motor Imitation Scale: performance measure of object and gesture imitation Unstructured Imitation Assessment	Ingersoll 2012
		Socially engaged imitation: observed during examiner/child play session	Landa 2011
	Participant behaviour: observation	Social Skills Observation: two x 5 minutes, during recess or free time in school	Hopkins 2011
		Joint attention and joint engagement during teacher-child or therapist-child play	Kaale 2012 Kim 2009
		Joint attention and joint engagement during mother-child play	Kaale 2012 Kasari 2006 Kasari 2010 Schertz 2013
		Structured Play Assessment	Goods 2013 Wong 2013
		Symbolic Play Test	Wong 2010
		Social skills during classroom observation	Goods 2013 Wong 2013
	Participant behaviour: report	Index of Empathy for Children and Adolescents: self-report Children's Social Behaviour Questionnaire (CSBQ): parent report	Begeer 2011
		ToM Questionnaire: teacher report	Fisher 2005
		Social Skills Rating System: parent report	Hopkins 2011
		KidScreen, parent-report quality of life measure	Baghdadli 2013
		Problem Behaviour Logs: parent report	Solomon 2004

Table 2. Outcome measures used (Continued)

		Ritvo-Freeman Real Life Rating Scale (RFRLRS): parent report	Wong 2010
		PDD-BI social approach, rated by both parent and teacher	Kim 2009
		Vineland Adaptive Behaviour Scales (VABS)	Schertz 2013
	Participant behaviour: direct assessment	Reynell Developmental Language Scales (post-test; six-month follow-up, 12-month follow-up)	Kasari 2006
		Mullen Scales of Early Learning (MSEL)	Landa 2011 Schertz 2013
		Expressive vocabulary test (five-year follow-up)	Kasari 2006
		Differential Abilities Scale (five-year follow-up)	Kasari 2006
		Test of Problem Solving (executive function)	Solomon 2004
	Acceptability of Intervention	Parent Adherence to Treatment & Competence: parent report Caregiver Quality of Involvement Scale: observational measure during parent-child play	Kasari 2010
		Teacher acceptability of intervention report	Wong 2013
		Children's Depression Inventory: self-report Beck Depression Inventory (BDI) (to assess parent depression) self-report	Solomon 2004
		Parenting Stress Index short form	Wong 2010
	Economic data	None	

Unlike the 'Summary of findings' table and the discussion of intervention effects in the main text, the principal organising element for this table is the methodology of each outcome measure. This underscores the great difficulty in comparing findings across studies due to wide variety in assessment scoring systems.

APPENDICES

Appendix I. Search strategies

CENTRAL 2013, Issue 7, last searched on 6 August 2013 (35 records)

Previous searches

CENTRAL 2012 (6), 10 July 2012 (126 records)

CENTRAL 2010 (3), 2 July 2010 (939 records)

#1 MeSH descriptor Child Development Disorders, Pervasive explode all trees

#2 autism*

#3 asperger*

#4 kanner*

#5 childhood schizophrenia

#6 pervasive developmental disorder*

#7 "PDD"

#8 language near/3 delay*

#9 speech near/3 disorder*

#10 (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9)

Ovid MEDLINE(R), 1946 to July Week 4 2013, last searched on 8 August 2013 (701 records)

Previous searches:

Ovid MEDLINE(R), 1946 to June Week 4 2012, 9 July 2012 (962 records)

Ovid MEDLINE(R) 1950 to July Week 1 2010, 19 July 2010 (4553 records)

1 exp child development disorders, pervasive/

2 autism\$.tw.

3 pervasive developmental disorder\$.tw.

4 pdd.tw.

5 (language adj3 delay\$).tw.

6 (speech adj3 disorder\$).tw.

7 childhood schizophrenia.tw.

8 kanner\$.tw.

9 asperger\$.tw.

10 or/1-9

11 randomized controlled trial.pt.

12 controlled clinical trial.pt.

13 randomi#ed.ab.

14 placebo\$.ab.

15 drug therapy.fs.

16 randomly.ab.

17 trial.ab.

18 groups.ab.

19 or/11-18

20 exp animals/ not humans.sh.

21 19 not 20

22 10 and 21

EMBASE (Ovid), 1980 to 2013 Week 31, last searched 6 August 2013 (992 records)

Previous searches:

Embase, 1980 to 2012 Week 27, 9 July 2012 (1979 records)

Embase, 1980 to 2010 Week 28, 19 July 2010 (3922 records)

1 exp autism/

2 autism\$.tw.

3 asperger\$.tw.

4 kanner\$.tw.
 5 childhood schizophrenia.tw.
 6 pervasive developmental disorder\$.tw.
 7 PDD.tw.
 8 language adj3 delay\$.tw.
 9 speech adj3 disorder\$.tw.
 10 or/1-9
 11 Clinical trial/
 12 Randomized controlled trial/
 13 Randomization/
 14 Single blind procedure/
 15 Double blind procedure/
 16 Crossover procedure/
 17 Placebo/
 18 Randomi#ed.tw.
 19 RCT.tw.
 20 (random\$ adj3 (allocat\$ or assign\$)).tw.
 21 randomly.ab.
 22 groups.ab.
 23 trial.ab.
 24 ((singl\$ or doubl\$ or trebl\$ or tripl\$) adj3 (blind\$ or mask\$)).tw.
 25 Placebo\$.tw.
 26 Prospective study/
 27 (crossover or cross-over).tw.
 28 prospective.tw.
 29 or/11-28
 30 10 and 29

CINAHLPlus (EBSCOhost) last searched 6 August 2013 (214 records)

Previous searches:

CINAHL 1937 to current, 9 July 2012 (211 records)

CINAHL 1937 to current, 23 July 2010 (847 records)

S26 S15 and 25

S25 S16 or S17 or S18 or S19 or S20 or S21 or S22 or S23 or S24

S24 TI(speech N3 disorder*) OR AB(speech N3 disorder

S23 TI(language N3 delay*) OR AB(language N3 delay*)

S22 TI(childhood schizophrenia*) or AB(childhood schizophrenia)

S21 TI(kanner*) or AB(kanner*)

S20 TI(asperger*) or AB(asperger*)

S19 TI(autis*) or AB(autis

S18 TI("PDD") or AB("PDD")

S17 TI (pervasive developmental disorder*) or AB (pervasive developmental disorder*)

S16 (MH "Child Development Disorders, Pervasive+")

S15 S1 or S2 or S3 or S4 or S5 or S6 or S7 or S8 or S9 or S10 or S11 or

S12 or S13 or S14

S14 TI(effectiv* study or effectiv* research) or AB(effectiv* study or

effectiv* research)

S13 "cross over*"

S12 crossover*

S11 (MH "Crossover Design")

S10 (tripl* N3 mask*) or (tripl* N3 blind*)

S9 (trebl* N3 mask*) or (trebl* N3 blind*)

S8 (doubl* N3 mask*) or (doubl* N3 blind*)
S7 (singl* N3 mask*) or (singl* N3 blind)
S6 (clinic* N3 trial*) or (control* N3 trial*)
S5 (random* N3 allocat*) or (random* N3 assign*)
S4 randomis* or randomiz*
S3 (MH "Meta Analysis")
S2 (MH "Clinical Trials+")
S1 MH random assignment

PsycINFO (Ovid) 1806 to July Week 5 2013 , last searched 6 August 2013 (284 records)

Previous searches:

PsycINFO 1806 to July Week 1 2012, searched 10 July 2012 (1361 records)

PsycINFO searched 2010 via EBSCOhost

1 autism/ or pervasive developmental disorders/ or aspergers syndrome/

2 Autistic Thinking/

3 pervasive developmental disorder\$.tw.

4 "pdd".tw.

5 autis\$.tw.

6 asperger\$.tw.

7 kanner\$.tw.

8 childhood schizophren\$.tw.

9 (language adj3 delay\$.tw.

10 (speech adj3 disorder\$.tw.

11 or/1-10

12 clinical trials/

13 (randomis* or randomiz*).tw.

14 (random\$ adj3 (allocat\$ or assign\$)).tw.

15 ((clinic\$ or control\$) adj trial\$).tw.

16 ((singl\$ or doubl\$ or trebl\$ or tripl\$) adj3 (blind\$ or mask\$)).tw.

17 (crossover\$ or "cross over\$").tw.

18 random sampling/

19 Experiment Controls/

20 Placebo/

21 placebo\$.tw.

22 exp program evaluation/

23 treatment effectiveness evaluation/

24 ((effectiveness or evaluat\$) adj3 (stud\$ or research\$)).tw.

25 or/12-24

26 11 and 25

PsycINFO (EBSCOhost), searched 23 July 2010 (891 records)

S29 S17 and S28

S28 S18 or S19 or S20 or S21 or S22 or S23 or S24 or S25 or S26 or S27

S27 TI(speech N3 disorder*) OR AB(speech N3 disorder*)

S26 TI(language N3 delay*) OR AB(language N3 delay)

S25 TI(childhood schizophrenia*) or AB(childhood schizophrenia*)

S24 TI(kanner*) or AB(kanner*)

S23 TI(asperger*) or AB(asperger*)

S22 TI(autis*) or AB(autis*)

S21 TI("PDD") or AB("PDD")

S20 TI (pervasive developmental disorder*) or AB (pervasive developmental disorder*)

S19 DE "Autistic Thinking
 S18 DE "Pervasive Developmental Disorders" OR DE "Aspergers Syndrome" OR
 DE "Autism" OR DE "Rett Syndrome" Search modes - Boolean/Phrase Interface
 S17 S1 or S2 or S3 or S4 or S5 or S6 or S7 or S8 or S9 or S10 or S11 or
 S12 or S13 or S14 or S15 or S16
 S16 TI (effectiv* research or effectiv* study) or AB (effectiv* research
 or effectiv* study)
 S15 DE "Meta Analysis"
 S14 TI((clinic* N3 trial*) or (control* N3 trial*)) OR AB ((clinic* N3
 trial*) or (control* N3 trial*))
 Database - PsycINFO Display
 S12 TI ((tripl* N3 mask*) or (tripl* N3 blind*)) or AB ((tripl* N3 mask*)
 or (tripl* N3 blind*))
 S11 TI((trebl* N3 mask*) or (trebl* N3 blind*)) or AB((trebl* N3 mask*) or
 (trebl* N3 blind*))
 S10 TI((doubl* N3 mask*) or (doubl* N3 blind*)) or AB ((doubl* N3 mask*)
 or (doubl* N3 blind*))
 S9 TI ((singl* N3 mask*) or (singl* N3 blind*)) or AB ((singl* N3 mask*)
 or (singl* N3 blind*))
 S8 TI((random* N3 allocat*) or (random* N3 assign*)) or AB((random* N3
 allocat*) or (random* N3 assign*))
 S7 TI (randomis* or randomiz*) or AB (randomis* or randomiz*)
 S6 DE "Placebo"
 S5 DE "Mental Health Program Evaluation"
 S4 DE "Psychotherapeutic Outcomes"
 S3 DE "Treatment Outcomes"
 S2 DE "Treatment Effectiveness Evaluation
 S1 DE "Random Sampling"

ASSIA (CSA) 1987 to current . Last searched 25 July 2010 (1603 records)

Not available in July 2012 or August 2013

((kw= (randomised or randomized or randomly or random or clinical trial* or control* or placebo* or group* or effectiveness
 within 3 research* or evaluat* witin 3 research*)) and((kw=(pervasive developmental disorder* or asperger* or autistic* or "pdd"
 or kanner* or childhood schizophrenia or language within 3 delay* or speech within 3 disorder*)))

Social Services Abstracts(Proquest), 1979 to current, last searched 6 August 2013 (59 records)

Previously searched in 2010 via CSA

(SU.EXACT("Autism") OR ("pervasive developmental disorder*" OR asperger* OR autistic* OR "pdd" OR kanner* OR "childhood
 schizophrenia" OR language NEAR/4 3 delay* OR speech NEAR/3 disorder*)) AND (SU.EXACT("Random Samples") OR (ran-
 domised OR randomized OR randomly OR random OR clinical trial* OR control* OR placebo* OR group* OR effectiveness NEAR/
 3 research* OR evaluat* NEAR/3 research*))

Social Services Abstracts(CSA), 1979 to current, 25 July 2010 (42 records)

((kw= (randomised or randomized or randomly or random or clinical trial* or control* or placebo* or group* or effectiveness
 within 3 research* or evaluat* within 3 research*)) and((kw=(pervasive developmental disorder* or asperger* or autistic* or "pdd"
 or kanner* or childhood schizophrenia or language within 3 delay* or speech within 3 disorder*)))

ERIC (Proquest), 1966 to current, last searched 6 August 2013 (284 records)

ERIC 1966 to current, searched 19 July 2012 (461 records)

Previously searched in 2010 via Dialog Datastar

SU.EXACT.EXPLODE("Asperger Syndrome" OR "Autism" OR "Pervasive Developmental Disorders") OR "PERVASIVE DEVEL-
 OPMENT*

DISORDER*" OR "PDD" OR speech NEAR/3 disorder* OR language NEAR/3 delay* OR autistic* OR asperger* OR kanner* OR
 "childhood

schizophren*) AND (SU.EXACT("Experimental Groups") OR SU.EXACT("Control Groups") OR SU.EXACT("Followup Studies") OR SU.EXACT("Longitudinal Studies") OR SU.EXACT("Program Effectiveness") OR ((prospective* OR "follow up" OR evaluat* OR compar* OR blind*) NEAR/3 stud*) OR (evaluat* NEAR/3 research*) OR ((compar* OR control*) NEAR/3 group*) OR random* OR intervention* OR experiment* OR trial*)

ERIC (Dialog Datastar) 1966 to current, searched 21 July 2010 (2982 records)

- 1.SEARCH:Pervasive-Developmental-Disorders#.DE.
- 2.SEARCH:pervasive ADJ developmental ADJ disorder\$
- 3.SEARCH:pdd
- 4.SEARCH:speech NEAR disorder\$
- 5.SEARCH:language NEAR delay\$
- 6.SEARCH:autis\$
- 7.SEARCH:asperger\$
- 8.SEARCH:kanner\$
- 9.SEARCH:childhood ADJ schizophrenia
- 10.SEARCH:1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9
- 11.SEARCH:Control-Groups.DE.
- 12.SEARCH:EXPERIMENTAL-GROUPS.DE.
- 13.SEARCH:LONGITUDINAL-STUDIES.DE.
- 14.SEARCH:FOLLOW-UP-STUDIES.DE.
- 15.SEARCH:PROGRAM-EFFECTIVENESS.DE.
- 16.SEARCH:((prospective\$ OR follow ADJ up OR evaluat\$ OR compar\$ OR blind\$) NEAR study).TI,AB.
- 17.SEARCH:(evaluat\$ NEAR research\$).TI,AB.
- 18.SEARCH:((compar\$ OR control\$) NEAR group\$).TI,AB.
- 19.SEARCH:random\$.TI,AB.
- 20.SEARCH:intervention\$.TI,AB.
- 21.SEARCH:experiment\$.TI,AB.
- 22.SEARCH:trial\$.TI,AB.
- 23.SEARCH:11 OR 12 OR 13 OR 14 OR 15 OR 16 OR 17 OR 18 OR 19 OR 20 OR 21 OR 22
- 24.SEARCH:10 AND 23

AutismData (<http://www.autism.org.uk/autismdata/>) last searched 7 August 2013 (59 records)

Searched in 2013 using the phrase "theory of mind"

Previous searches:

10 July 2012 (151 records)

27 July 2010 (29 records)

Searched in 2010 and 2012 using the keywords **random* or RCT or control* or trial***

ICTRP searched 7 August 2013 (85 records)

Previous searches:

10 July 2012 (234 records)

Simple search using terms **autis* OR asperg* OR pervasive developmental disorder***

metaRegister of Controlled Trials (mRCT) searched 7 August 2013 (19 records)

Searched in 2013 using the phrase "theory of mind"

Previous searches:

25 July 2010 (297 records)

autis* OR asperg* OR pervasive developmental disorder*

CONTRIBUTIONS OF AUTHORS

Draft the protocol	S Fletcher-Watson
Develop a search strategy	S Fletcher-Watson
Select which trials to include (2 people + 1 arbiter in the event of dispute)	S Fletcher-Watson, E Manola, F McConnell Arbiter: H McConachie
Extract data from trials (2 people)	S Fletcher-Watson, E Manola, F McConnell
Enter data into RevMan (Cochrane software)	S Fletcher-Watson
Carry out the analysis	S Fletcher-Watson
Interpret the analysis	S Fletcher-Watson, H McConachie
Draft the final review	S Fletcher-Watson, H McConachie
Keep the review up to date	S Fletcher-Watson

DECLARATIONS OF INTEREST

A £1000 grant was provided by the charity Research Autism, to pay for some research assistance.

Sue Fletcher-Watson and Helen McConachie - both of these authors have been involved in the development of a therapeutic iPad app for preschoolers with autism, targeting skills relevant to the current review. An RCT of this app has recently been completed and will be published in due course. This study may be deemed eligible for inclusion in a future version of this review. The app is now available on the commercial market and the authors could both receive royalties from sales of the app if these exceed a certain threshold.

Fiona McConnell - none known.

Eirini Manola - none known.

SOURCES OF SUPPORT

Internal sources

- No sources of support supplied

External sources

- Research Autism, UK.

Funding for training and to employ a research associate

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

Influenced by changes to diagnostic criteria in DSM-5 ([APA 2013](#)), and the direction of theory which suggests Theory of Mind (ToM) primarily underlies social and communication impairments in autism, we removed the domain of FLEXIBILITY AND IMAGINATION from the expected Primary Outcomes.

A further criterion under “Types of Interventions” was implemented to exclude complex, broad-based interventions addressing an array of social and communication behaviours. While the authors note the efficacy and appropriateness of this approach, these interventions need to be excluded from our review because they cannot be described to relate directly to the ToM model of autism.