

Case Study 1: An Evidence-Based Practice Review Report

How effective are Theory of Mind (ToM) based group interventions on improving emotion recognition skills of children with Autism Spectrum Disorder (ASD)?

Summary

The effectiveness of Theory of Mind (ToM) based group interventions on improving emotion recognition skills of children with Autism Spectrum Disorder (ASD) is examined in this systematic literature review. ToM based group interventions are interventions that teach a range of precursor or component skills of ToM using a range of group activities. The focused outcome of this review is the precursor skill of emotion recognition.

A systematic literature search was performed in January 2018 using the databases ERIC, PsycINFO and Medline. A total of six peer-reviewed journal articles were generated at the end of the inclusion and exclusion process. The relevance of these studies to the review question was evaluated using the weight of evidence framework (Gough, 2007). The studies were rated according to their methodological quality, appropriateness of design and topic relevance. Most of the studies found mixed outcomes, as most of the studies that found large or medium effect sizes also showed minimal effects on other relevant outcomes. To conclude, a solid evidence base for the effectiveness of the target intervention on improving emotion recognition skills of children with ASD has yet to be established. Recommendations for further research are discussed.

Introduction

In the UK, autism spectrum disorder (ASD) remains one of the most common primary types of special educational needs among all students with Education, Health and Care plans (EHC) or statements (Department for Education, 2017).

Autism spectrum disorder (ASD) is a form of neurodevelopmental disorder characterised by persistent social and communication difficulties as well as repetitive behaviour across multiple contexts (American Psychiatric Association, 2013); individuals with ASD were found to have difficulties such as developing appropriate social relationships, interpreting social situations and sharing emotions with others (American Psychiatric Association, 2013; Charman & Stone, 2008; Stichter, O'Connor, Herzog, Lierheimer & McGhee, 2012).

The Theory of Mind (ToM) model can be used to understand the social and communication difficulties associated with ASD. Theory of Mind (ToM) refers to the ability to understand others' mental states (Premack & Woodruff, 1978), such as intentions, emotions and desires (Wellman, Cross & Watson, 2001), which allows individuals to infer on others' beliefs, predict how others might behave and generate appropriate responses in social situations (Baron-Cohen et al., 1985).

Research showed that 80% of typically developing children develop the understanding of diverse desire, diverse belief, knowledge access, false belief and hidden emotions in a sequential order (Peterson & Wellman, 2009; Wellman, Fang & Peterson, 2011). There were subtle cultural difference in the developmental sequence, such that the development of knowledge access and diverse belief were found to be reversed in Iranian and Chinese children as compared to children in the US or western countries (Shahaeian, Peterson, Slaughter & Wellman, 2011; Wellman, Fang, Liu, Zhu & Liu, 2006; Wellman et al., 2011).

Research proposed that such difference was due to the cultural emphasis on

knowledge acquisition and consensual learning in Iran and China, as opposed to an emphasis on critical thinking in the US or western countries. The importance of environmental influence on ToM development can also be interpreted to understand the social communication deficits shown by individuals with ASD.

Harms, Martin and Wallace (2010) suggested that since children with ASD lack social orientation from infancy, they would have less social experience to facilitate the development of ToM, which led to difficulties associated with social interactions.

When using the ToM model to understand social and communication difficulties experienced or displayed by individuals with ASD, it is also important to investigate the precursor or component skills of ToM associated with such difficulties.

Although recent fMRI studies have suggested that there could be two aspects of ToM (cognitive and affective) that were differentially impaired in individuals with ASD (Kim et al., 2016), evidence showed that emotion recognition, as one of the precursor skills associated with ToM development (Fletcher-Watson, McConnell, Manola, & McConachie, 2014), is commonly impaired in individuals with ASD. It was found that children with ASD encountered difficulties reporting their own emotions (Capps, Yirmiya & Sigman, 1992), recognising basic facial expressions, understanding complex emotions of others (Kuusikko et al., 2009) and matching facial expression to emotions (Celani et al., 1999). Although the above evidence suggests that emotion recognition skills are impaired in individuals with ASD, other studies also found that emotion understanding can be improved through an increased exposure to social situations and to a variety of facial expressions (Bajgar, Ciarrochi, Lane & Deane, 2005; Harms et al., 2010). The current review will focus on ToM based interventions that aim to improve children's emotion

recognition through increased exposure to social situations (i.e. group setting) and explicit teaching of emotion recognition skills.

ToM based interventions aimed at explicitly teaching individuals with ASD to identify mental states of self and others through targeting the precursor or component skills of ToM (Fletcher-Watson et al., 2014; Swettenham, 2000). These interventions are different to interventions that mainly use behavioural techniques, such as reward and punishment, to reinforce desirable social behaviour. ToM interventions that target emotion recognition skills can be implemented in the forms of home and/or clinic-based individual training (Bell & Kirby, 2002; Bernad-Ripoll, 2007; Hadwin et al., 1997; Williams, Gray & Tonge, 2012), mixed individual and group training in a school (Feng et al., 2008) as well as setting (school/ clinic) based small group training (Begeer et al., 2011; 2015; Gevers, Clifford, Mager & Boer, 2006; Solomon, Goodlin-Jones & Anders, 2004; Soorya et al., 2015; Stichter et al., 2012; Stichter, Herzog, Owens & Malugen, 2016; Waugh & Peskin, 2015). Only a limited number of studies on the relevant interventions were set in a school (Bauminger, 2007; Bauminger-Zviely, Eden, Zancanaro, Weiss & Gal, 2013; Feng et al., 2008; Stichter et al., 2016), as most of the other studies were set in clinics or treatment centres. The ToM interventions can be delivered through different means, such as computer games (Bauminger-Zviely et al, 2013; Silver & Oakes, 2001; Rice, Wall, Fogel & Shic, 2015), watching emotion recognition focused videos (Williams, Gray & Tonge, 2012), use of thought bubble (Waugh & Peskin, 2015; Wellman et al., 2002), social stories and self-as-model videos (Bernad-Ripoll, 2007), or group activities. The structure of group interventions and the variety of group activities will be studied in greater detail later on in this review of literature.

It is important to review the evidence related to the effectiveness of ToM based group interventions on improving emotion recognition skills of children with ASD, as it was stated in the SEND Code of Practice (DfE & DH, 2015) that educational psychologists (EPs), as one of the external service providers, should work closely with school staff to ensure that interventions are based on reliable evidence of effectiveness and would address the learning needs of students. In order to address the specific social communication needs of children with ASD, EPs should review the evidence base of ToM interventions and make appropriate advice on intervention implementation.

This systematic literature review aims at addressing the following question:

How effective are Theory of Mind (ToM) based group interventions on improving emotion recognition skills of children with ASD?

Critical Review of the Evidence Base

Systematic Literature Search

In January 2018, the online databases ERIC (Education Resources Information Center), Medline and PsycINFO were used to perform a systematic literature search on the review topic. The databases were accessed through the OVID and EBSCOHost platforms. The combinations of search terms used with each database are shown in Table 1.

Table 1
Combination of search terms used with online databases

	Theory of Mind		Intervention		ASD		Children
ERIC	theory of mind or tom	AND	intervention or training or teaching	AND	autism or asd or autism spectrum disorder or autistic	AND	Education level: Early Childhood Education, Elementary Education
PsycINFO	"Theory of mind".mp		TRAINING/ or TEACHING/ or GROUP		Autism Spectrum Disorder.mp or		Childhood

	or "Theory of mind"/	INTERVENTION/ or SCHOOL BASED INTERVENTION/ or INTERVENTION/	Autis*.mp or ASD.mp or Autism Spectrum Disorder/	
Medline	"theory of mind".mp or "theory of mind"/	"intervention or training or teaching".mp or "Early Intervention (Education)"/ or 'school based intervention'.mp	'ASD'.mp or "Autistic Disorder".mp or Autism Spectrum disorder/ or "Autism Spectrum Disorder".mp	"newborn infant (birth to 1 month)" or "infant (1 to 23 months)" or "preschool child (2 to 5 years)" or "child (6 to 12 years)"

The literature search and selection processes are illustrated in Figure 1. A total of 101 peer-reviewed studies written in English were generated as a result of the initial search. After removing duplicates, 77 studies were reviewed at a title and abstract level with the purpose of filtering out studies irrelevant to ToM based interventions for children with ASD. Full-text screening was first performed on 34 studies to exclude interventions that did not address emotion recognition skills, out of which 17 were screened again at a full-text level. The inclusion and exclusion criteria are listed in Table 2; the full references of the six studies included in this review are listed in Table 3. See Appendix A for the list of excluded studies.

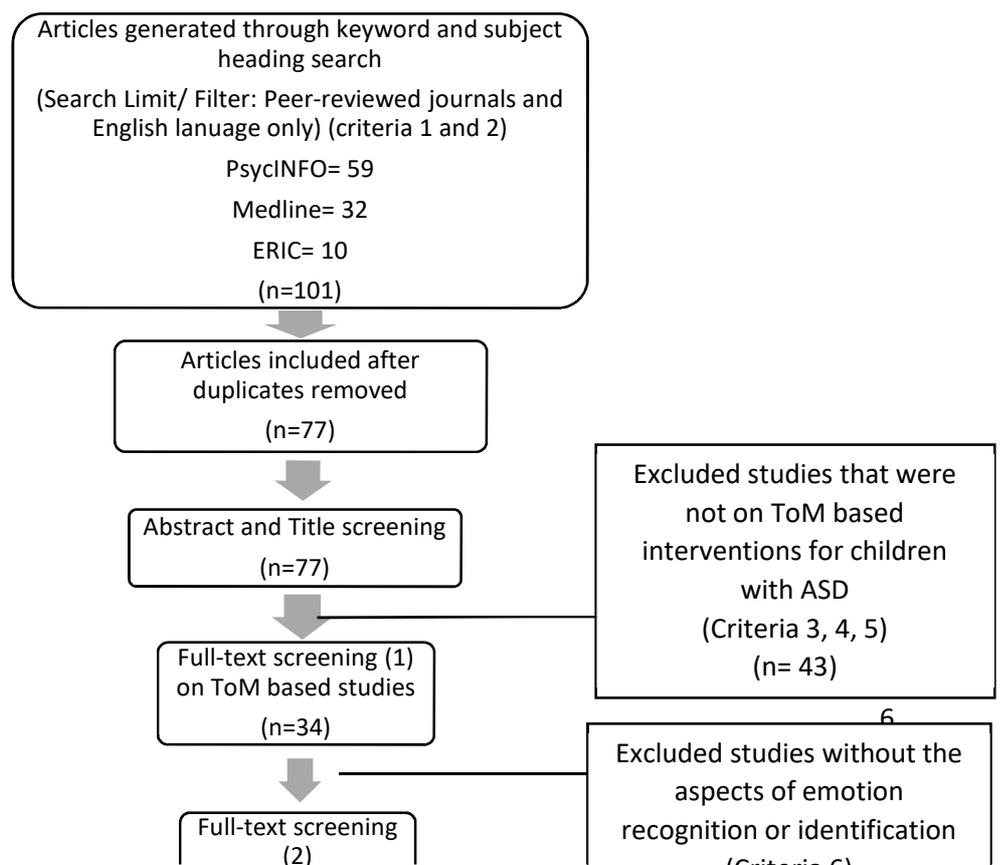


Figure 1. Literature search and selection process flowchart.

Table 2

Inclusion and exclusion criteria

	Inclusion criteria	Exclusion criteria	Rationale
1. Type of publication	The study must be published in a peer-reviewed journal.	Studies from sources other than peer-reviewed journals.	Peer-reviewed publications are of a higher standard, as they have been evaluated by expert reviewers.
2. Language of publication	The entire study must be published in English.	The whole study or parts of the study are not published in English.	This is to ensure that the reviewer is able to understand the full study.
3. Participants	The study must include participants with Autism Spectrum Disorder (ASD) (DSM-5), or the DSM-IV equivalent of Asperger's Disorder, High Functioning Autism, Autistic Disorder and Pervasive Developmental Disorder, Not Otherwise Specified (PDD-NOS).	The study only sampled participants with diagnoses other than ASD (e.g. ADHD, Rett's Disorder and Childhood Disintegrative Disorder).	The current review focuses on interventions for participants with ASD.

4. Intervention based	The study must include at least one teaching, intervention or training phase.	The study contained no teaching, intervention or training element (e.g. assessment focused).	The current review focuses on the effect of an intervention, teaching or training, instead of assessments.
5. Focus and outcome of intervention (1)	The study must measure the direct effect(s) of at least one ToM based intervention.	The study measures the effect of other mediating factors such as siblings and parental style.	The current review focuses on the direct effect of ToM interventions.
6. Focus and outcome of intervention (2)- emotion recognition	Emotion knowledge, recognition, identification or awareness must be explicitly taught in the intervention.	The intervention covered other components of ToM (e.g. false belief, perspective taking), instead of emotion recognition.	The current review focuses on the learning of the precursor skill of ToM, which is emotions recognition.

	Inclusion criteria	Exclusion criteria	Rationale
7. Outcome of intervention	The study must measure emotion recognition related outcomes and report statistical tests used.	The study did not measure emotion recognition related outcomes or/and report statistical tests used.	The effect of an intervention on emotion recognition skills had to be measured and analysed using statistical test(s).
8. Level of intervention	Intervention must be administered in a small group setting throughout the whole period (4-7 participants per group).	The participants received one-on-one intervention throughout (e.g. computer-based individual) or in separate phases of an intervention (e.g. group phase followed by individual teaching phase or vice versa).	The current review focuses on interventions that can be administered predominantly in a small group format.
9. Setting of intervention	The whole intervention must be conducted in settings outside of the home.	The whole or parts of the intervention were conducted in a home environment.	The current review focuses on setting-based interventions that can potentially be applied in schools.

Table 3

Studies included in the current review

1. Begeer, S., Gevers, C., Clifford, P., Verhoeve, M., Kat, K., Hoddenbach, E., & Boer, F. (2011). ToM training in children with autism: A randomized controlled trial. *Journal of Autism and Developmental Disorders*, 41(8), 997–1006.
2. Begeer, S., Howlin, P., Hoddenbach, E., Clauser, C., Lindauer, R., Clifford, P, Gevers, C., Boer, F., & Koot, H. M. (2015). Effects and Moderators of a Short ToM Intervention for Children with Autism Spectrum Disorder: A Randomized Controlled Trial. *Autism Research : Official Journal of the International Society for Autism Research*, 8(6), 738–48.
3. Solomon, M., Goodlin-Jones, B. L., & Anders, T. F. (2004). A Social Adjustment Enhancement Intervention for High Functioning Autism, Asperger’s Syndrome, and Pervasive Developmental Disorder NOS. *Journal of Autism and Developmental Disorders*, 34(6), 649–668.
4. Soorya, L. V., Siper, P. M., Beck, T., Soffes, S., Halpern, D., Gorenstein, M., Kolevzon, A., Buxbaum, J., Wang, A. T. (2015). Randomized comparative trial of a social cognitive skills group for children with autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, 54(3), 208–216.
5. Stichter, J. P., O’Connor, K. V., Herzog, M. J., Lierheimer, K., & McGhee, S. D. (2012). Social competence intervention for elementary students with Aspergers Syndrome and high functioning autism. *Journal of Autism and Developmental Disorders*, 42(3), 354–366.

6. Stichter, J. P., Herzog, M. J., Owens, S. A., & Malugen, E. (2016). Manualization, feasibility, and effectiveness of the school-based Social Competence Intervention for Adolescents (SCI-A). *Psychology in the Schools*, 53(6), 583–600.
-

Mapping the Field

The systematic literature search has identified six studies that are relevant to the review topic. The details of the studies are outlined in Table 4.

Critical Evaluation

Gough's (2007) weight of evidence framework was used to evaluate the identified studies, and separate judgement was made on three main aspects. The weight of evidence A (WoE A) is the judgement on the quality of a study, which was carried out using a coding protocol from Gersten et al. (2005)'s quality indicator for group experimental and quasi-experimental research. The weight of evidence B (WoE B) judgement on the appropriateness of study design was made using the typology of evidence (Petticrew & Roberts, 2003), which outlined the types of research design that were least to most suitable for investigating the effectiveness of interventions. The weight of evidence C (WoE C) judgement was based on topic relevance; the intervention setting, training and supervision provided to intervention administrators and curriculum relevance were analysed. An example of the WoE A coding protocol can be found in Appendix C; the judgement criteria for WoE B and C as well as the ratings for WoE A, B and C can be found in Appendix B (Table B1-6). An overall judgement, the weight of evidence D (WoE D), was formulated by averaging the ratings of WoE A, B and C, which represents how relevant a study is to the review question (Gough, 2007). For the purpose of this review, a WoE D rating of 1.67 or below is considered as low, 1.68 to 2.32 is medium and

above 2.32 is high. A summary table of the weight of evidence is shown below (Table 5).

Table 4

Mapping the field

Study	Country	Setting; Intervention group size	Total sample (n, diagnosis)	Age (years)	Name of intervention	Intervention (duration, lesson structure)	Number of sessions on emotional understanding	Administrator; training or support given to administrator
Begeer et al. (2011).	Netherlands	Unspecified setting; 5-6 per group	n= 40, Diagnosis of ASD	8-13	Manualised ToM training (programme developed by Steerneman, 1994; Steerneman et al., 1996)	16 weekly 1.5 hour sessions. Lesson structure: not specified in the study. All end with 15 mins group training with parents at the end.	n/a; not specified	Unclear description of therapists. Supervision: Therapists were trained following a manual and provided with on- going supervision
Begeer et al. (2015)	Netherlands	Child Psychiatric Centre; 5-6 per group	n= 97, Diagnosis of ASD	7-12	Manualised Mini ToM training (short version of Manualised ToM training, adapted from programme developed by Steerneman, 1994; Steerneman et al., 1996)	8 weekly 1 hour sessions. A short version of ToM intervention. Lesson structure: homework discussion (recap of the previous lesson), games and exercise related to the theme (teaching), children summarise the session to parents, explanations of	2/ 8 sessions	Unclear who administered the intervention (workers at a child psychiatric centre). Supervised by certified therapists, licensed counselling psychologists who were specifically trained for the intervention.

homework for the following week (published in Hoddenbach et al., 2012).

Study	Country	Setting; Intervention group size	Total sample (n, diagnosis)	Age (years)	Name of intervention	Intervention (duration, lesson structure)	Number of sessions on emotional understanding	Administrator; training or support given to administrator
Solomon et al. (2004)	USA	university medical centre; 4-5 per group	n=18, Diagnosis of high functioning HFA, Asperger Syndrome, PDD-NOS.	8-12	Social adjustment enhancement curriculum	20 weekly 1.5 hour sessions. Two 10 week modules. Lesson structure: welcome song, check-in time for discussion of previous week's homework, snack time (practice conversation skills), lesson time, motor activity time and joke time.	6/ 18 teaching sessions (All in module 1)	Group leaders: Psychologists, a psychiatrist, a speech and language pathologist (therapist), one assistant (videotaped activities and acted as role models). Supervision: group leaders met weekly to discuss issues in the group.

Soorya et al. (2015)	USA	Unspecified setting; 4-6 per group	n= 69, Diagnosis of ASD	8-11	Seaver-NETT (Nonverbal communication, emotion recognition, and ToM training)	12 weekly 1.5-hour sessions. Lesson structure: 15 minutes free-play/ snacks, 60 minutes instruction, 15 minutes circle time.	3/11 teaching sessions	Group leaders: clinical psychologists. Assisted by specifically trained therapy assistants. Weekly (within group) supervision for each treatment team by the lead therapist.
Study	Country	Setting; Intervention group size	Total sample (n, diagnosis)	Age (years)	Name of intervention	Intervention (duration, lesson structure)	Number of sessions on emotional understanding	Administrator; training or support given to administrator
Stichter et al. (2012)	USA	Classrooms in treatment centre for autism and neuro-developmental disorder,; 4-7 per group	n= 20, Diagnosis of autism, Aspergers Syndrome, PDD-NOS and ASD	6-10	Social Competence Intervention - Elementary (SCI-E) program (adapted from the SCI-Adolescence programme for elementary school aged children)	20 twice weekly 1-hour sessions. Lesson structure: review, topic introduction, modelling and practice	8/ 20 sessions	Group leader: administrator with Master's degree, ASD specialisation and training in special education. No specific programme training or supervision mentioned

Stichter et al. (2016)	USA	Secondary (middle and junior high) school; 4-6 per group	n= 34, not all with ASD diagnosis (ADHD/ no diagnosis or other diagnoses)	11-15	Social Competence Intervention for Adolescence (SCI-A) (from Stichter et al., 2010)	2 to 3 times per week, 45 minutes per lesson. About 23 hours in total. Lesson structure: review, topic introduction, modelling and practice.	All sessions; author claimed that facial expression was being taught reinforced in all sessions, but a large proportion of sessions will target three or more skills.	Special education teachers and speech and language pathologist. Specifically trained and supervised by research staff throughout.
------------------------	-----	---	---	-------	---	---	---	--

Table 5

Weight of evidence rating summary

Study	Methodological quality (WoE A)	Appropriateness of design (WoE B)	Topic Relevance (WoE C)	Overall weight of evidence (WoE D)
Begeer et al. (2011).	1 (low)	3 (high)	1.67 (low)	1.89 (medium)
Begeer et al. (2015)	3 (high)	3 (high)	2 (medium)	2.67 (high)
Solomon et al. (2004)	1 (low)	2 (medium)	2.33 (high)	1.78 (medium)
Soorya et al. (2015)	3 (high)	3 (high)	2 (medium)	2.67 (high)
Stichter et al. (2012)	1 (low)	2 (medium)	2 (medium)	1.67 (low)
Stichter et al. (2016)	1 (low)	2 (medium)	3 (high)	2 (medium)

Participants

This review consists of a total of 278 participants. The age of the participants ranges from 6 to 15 years old. All participants included in this review had a full-scale or verbal IQ score of at least 70. UK samples were not included in this review. The participants from two of the studies were from the Netherlands (Begeer et al., 2011; 2015), whereas participants from the other studies (Solomon et al., 2004; Soorya et al., 2015; Stichter et al., 2012; 2016) were from the United States.

Convenience sampling was used to recruit participants in all studies. The participants were recruited from an academic centre for children and adolescent psychiatry (Begeer et al., 2011; 2015), websites (Solomon, 2004; Stichter, 2012), advertisements (Soorya, 2015; Stichter, 2012) or referrals from other professionals such as paediatrician, local practitioners, physician, psychiatrist, community service coordinators or school staff (Solomon et al., 2004; Soorya, 2015; Stichter, 2012; 2016). Such sampling method does not ensure that the sample is representative of the population, and sampling error cannot be estimated (Barker, Pistrang & Elliot, 2016).

The participants in four studies (Beeger et al., 2011; 2015; Solomon et al., 2004; Soorya et al., 2015) were diagnosed with autistic disorder, Asperger's disorder or pervasive developmental disorder-not otherwise (PDD-NOS) specified using the DSM-IV-TR, which was equivalent to the diagnosis of autism spectrum disorder (ASD) in the DSM-5 (American Psychiatric Association, 2000; 2013). In Stichter et al. (2012), the Autism Diagnostic Observation Schedule (ADOS; Lord et al. 2003) and Autism Diagnostic Interview-Revised (ADI-R; Lord et al. 1994) were used to

confirm that participants met the criteria for pervasive developmental disorders. Stichter et al. (2016) is the only study with non-ASD participants; about half of the sample (47.1%) were previously diagnosed with ASD, a small proportion (17.6%) was diagnosed with ADHD and about a quarter of participants (26.5%) were not formally diagnosed. A study would receive a higher WoE A rating if it has provided detailed information on the disabilities/ diagnosis of their participants.

Design

The information on study design is shown in Appendix B (Table B4). Four studies (Begeer et al., 2011; 2015; Solomon et al., 2004; Soorya et al., 2015) reported using randomisation to divide participants into treatment and control groups; details of the randomisation process were only reported in Begeer et al. (2015). Out of these four studies, three used the waitlist control method and one engaged the control group in facilitated play (controlled for the instructions given) (Soorya et al. 2015). Stichter et al. (2012; 2016) adopted a quasi-experimental (one-group pretest-posttest) design. The problems associated with this design, such as threats to internal validity and construct validity, meant that the effects measured cannot be directly attributed to the intervention (Barker, Pistrang & Elliot, 2016). Hence, a lower WoE rating was given to quasi-experimental studies.

Half of the studies reported the time when outcomes were measured; Soorya et al. (2015) and Beeger et al. (2015) assessed participants one-week post-intervention, with three and six months follow-up assessment respectively. Both studies received a high methodological relevance rating (WoE A). Stichter et al. (2012) carried out assessments two weeks post-intervention, and the intervention effects were not statistically significant. It was recommended by Gersten et al. (2005) that

data should be collected within a few days at the end of an intervention, as the results might be affected by a fading intervention effect, other uncontrolled events or subsequent instructions. This element is captured by the coding protocol (WoE A).

Measures

All studies used dependent variables relevant to emotion recognition, which is one of the inclusion criteria of the current review. A total of four previously developed assessment tools were used amongst the studies to capture the outcomes of the interventions. These measures were the Levels of Emotional Awareness Scale for Children (LEAS-C) (Bajgar, Ciarrochi, Lane & Deane, 2005), the Diagnostic Analysis of Non-Verbal Accuracy-2, Adult Facial Expressions (DANVA-2-AF) and Child Facial Expressions (DANVA 2-CF) (Nowicki & Carton, 1992) and the Reading the Eye in the Mind Test (RMET) (Baron-Cohen, Wheelwright & Hill, 2001). The LEAS-C is a questionnaire with 12 hypothetical scenarios; children were required to describe basic and complex emotions as well as engage in perspective taking tasks. Moderate internal consistency and acceptable convergent validity were reported (Begeer et al., 2011; 2015). The DANVA tests emotional identification using 24 photos of happy, sad, angry and fearful facial expressions in high or low intensity (i.e. more subtle and difficult to identify), with the DANVA-2-AF being pictures of adults' faces and DANVA 2-CF being children's faces. The Reading the Eye in the Mind Test (RMET) assesses emotion identification and thoughts inferencing using photos of the eye region. Only one study reported on test-retest reliability of the measure (DANVA2) (Solomon et al., 2004); information regarding the validity of the measures has not been reported.

All studies did not report whether or not test administrators had to obtain certain qualifications or undergo training to use the tests.

Intervention

The interventions were conducted in small groups of four to seven participants. The total intervention length ranged from eight to about 30 sessions (23 hours), the intensity ranged from one to three times a week and the duration from 45 minutes to 1.5 hours per session. All studies reported that the interventions were carried out following a manual or curriculum. The degree of emphasis put on the target element of ToM (emotion recognition) varied between the interventions. Emotion identification skills were being taught or maintained more intensively in some interventions (Solomon et al., 2004, Stichter et al., 2012; 2016), whereas it was only taught in about 25% of the sessions in some interventions (Begeer et al., 2015; Soorya et al., 2015). The curriculum relevance is reflected in WoE C.

Four studies started each intervention session by reviewing the content covered in the last session or homework set for the week (Begeer et al., 2015; Solomon et al., 2004; Stichter et al., 2012; 2016). The range of methods used to teach emotions included modelling of facial expressions and gestures, teaching of emotion words and analysing video, photo or real-life emotion-related scenarios. Most interventions were designed to have participants apply the learnt concept or skills through group activities (e.g. discussion, board game, role play, free play, facilitated play and group games) (Begeer et al., 2015; Solomon et al., 2004; Soorya et al., 2015; Stichter et al., 2012; 2016). Two studies elaborated on the activities, scaffolding strategies or instructions used (Solomon et al., 2004; Soorya et al., 2015); authors of the other studies claimed that such detail was available in

the intervention manual or related papers. A few studies also reported parental involvement (Begeer et al., 2011; 2015; Solomon et al., 2004), which was in the form of parent training or children summarising sessions to parents. Lesson plans were not provided by any studies.

Only one study was conducted in a school environment (Stichter et al., 2016), hence the high topic relevance (WoE C); other studies were conducted in clinics or unspecified settings. While studying the effectiveness of interventions in a clinic setting allows a higher degree of control over the set-up and higher internal validity, the external validity of clinical studies is often lower than field research that is conducted in a school setting (Barker, Pistrang & Elliot, 2016).

It is important that studies clearly describe their intervention administrator as well as the specific training and supervision that they received, as it provides vital information on the individuals capable of carrying out the intervention outside of the context of the study (Gersten et al., 2005). It was described in four studies that the group leaders were psychologists, psychiatrist, speech and language pathologist (therapist) and/or specialist teachers, with or without the support of specifically trained assistants. One to three adults were present in each small group. Begeer et al. (2011; 2015) did not provide detailed information on the background of the therapists administering the intervention. Except for Stichter et al. (2012), all studies specified that there were on-going or weekly supervision from professionals (research staff, psychologists, speech and language therapists) throughout the duration of the interventions. Implementation fidelity was monitored in three studies (Begeer et al., 2011; Soorya et al., 2015; Stichter et al., 2016) through videotape reviews, checklists and/or observation.

Outcome and Effect Sizes

Effect sizes of the outcomes related to emotional recognition were reported in most of the studies (Begeer et al., 2011; 2015, Soorya et al., 2015, Stichter et al., 2012; 2016). Effect sizes were calculated from available data in Solomon et al. (2004). Effect sizes reported in partial eta square (Begeer et al., 2015, Stichter et al., 2006) have been converted into the common effect size, Cohen's d , for comparison purpose. For interpretation purposes, Cohen's (1992) definition of small ($d= 0.2-0.49$), medium ($d= 0.5-0.79$) and large ($d \geq 0.8$) effect sizes were used. Effect sizes below 0.2 are said to be minimal. The corresponding outcomes, effect sizes and WoE D of each study can be found in Table 6.

While studies with high WoE D showed either medium (Soorya et al., 2015) or minimal effects (Begeer et al., 2015), the studies with medium WoE D all found large effect sizes with some of the outcomes (Begeer et al., 2011, Solomon et al., 2004 and Stichter et al., 2016). Minimal effects were often found alongside medium to large effect sizes (Begeer et al., 2011; Soorya et al., 2015; Stichter et al., 2016). Therefore, the effectiveness of the ToM interventions on emotion recognition in children with ASD is inconsistent. These results should be interpreted with caution when inferring the effectiveness of ToM based group interventions for children with ASD. The mixed results suggested that although the interventions might have medium to large effects on improving some aspects of emotional recognition, the effects might not have been generalised across all aspects of emotional recognition. Another possibility is that the improvements were not captured by the post-intervention measure.

Moreover, the relationship between effect size and sample size as well as related biases should not be overlooked. The largest effect size out of all studies was found in Solomon et al. (2004) that had a small sample size in both the treatment (n=9) and control group (n=9), which the author predicted that the study would lack a high level of statistical power. In this case, one should be aware of the possibility that there could be superrealisation bias in studies with small sample sizes (Cronbach et al., 1980; Slavin & Smith, 2009), such that participants received high-quality teaching and their progress was more carefully monitored in small sample size studies as opposed to studies with a bigger sample size. Qualitative investigation is needed to support this claim. On the other hand, Stichter et al. (2012) only had 20 participants in total and only included 8 participants in one measure, which lacked sufficient power to find a statistically significant effect (Cohen, 1992).

Table 6

Relevant measures, outcome and effect sizes of the included studies

Study	Relevant measure(s)	Outcome	Effect sizes		Findings	Effect size range (Cohen's <i>d</i>) and descriptors ^c	WoE D
			(n^2_p)	(Cohen's <i>d</i>)			
Begeer et al. (2011).	Levels of Emotional awareness scale (LEAS-C): Mixed, complex and total emotions	Total emotional awareness	--	0.38 ^b (d_{ppc2} , $p=0.37$)	No statistically significant difference in emotional awareness	0.38- 1.19 (small to large)	1.89 (medium)
		Mixed emotions understanding		0.84	Significant difference between control and treatment in mixed and complex emotions understanding		
		Complex emotions understanding		1.19			
Begeer et al. (2015)	Levels of Emotional awareness scale (LEAS-C): mixed, complex and total emotions	Total emotional awareness	0.03	0.30 ^b (d_{ppc2} , $p=0.07$)	No statistically significant differences in emotional understanding (all types)	0.09- 0.30 (minimal to small)	2.67 (high)
		Mixed emotions understanding	0.00	0.09 ^b (d_{ppc2} , $p=0.48$)			
		Complex emotions understanding	0.01	0.22 ^b (d_{ppc2} , $p=0.36$)			
Solomon et al. (2004)	Diagnostic Analysis of Non-Verbal Accuracy-2, Adult Facial Expressions (DANVA-2-AF) and Child Facial Expressions (DANVA 2-CF)	Facial expression recognition (total)	--	1.83 ^a	Significant improvement in total, child and adult facial expression identification	1.24- 1.83 (large)	1.78 (medium)
		Recognition of faces (adult faces)		1.32 ^a			
		Recognition of faces (child faces)		1.24 ^a			

Study	Relevant measure(s)	Outcome	Effect sizes		Findings	Effect size range (Cohen's <i>d</i>) and descriptors ^c	WoE D
			(n^2_p)	(Cohen's <i>d</i>)			
Soorya et al. (2015)	Combining RMET, DANVA and other tests	Facial expression recognition (total)	--	0.09 ^b	No statistically significant improvement on the social cognition composite score (DANVA and RMET were parts of the score) or in RMET.	0.09- 0.56 (minimal to medium)	2.67 (high)
	Reading the Eye in the Mind Test (RMET)	Recognising emotions from eye region photos		0.14 ^b			
	Diagnostic Analysis of Non-Verbal Accuracy2 (DANVA2)	Recognising low intensity emotions in child faces		0.56	Significant improvement in identifying low-intensity emotion in child faces.		
Stichter et al. (2012)	Diagnostic Analysis of Non-Verbal Accuracy2, Child Facial Expressions (DANVA-2-CF)	Recognition of faces (child faces)	--	0.07 (n=8)	No statistically significant improvement in emotional identification or labelling emotional or mental states.	0.07- 0.13 (minimal)	1.67 (low)
	Reading the Eye in the Mind Test (RMET)	Recognising emotions from eye region photos		0.13			

Study	Relevant measure(s)	Outcome	Effect sizes		Findings	Effect size range (Cohen's <i>d</i>) and descriptors ^c	WoE D
			(n^2_p)	(Cohen's <i>d</i>)			
Stichter et al. (2016)	Combining RMET and DANVA	Overall effect (combining both measures)	0.06	0.51 ^b ($p=0.398$)	Overall improvement in emotion recognition was not statistically significant.	-0.12- 1.14 (Minimal negative to large)	2 (medium)
	Reading the Eye in the Mind Test (RMET)	Recognising emotions from eye region photos		-0.12 ^b	No significant improvement		
	Diagnostic Analysis of Non-Verbal Accuracy2, Child Facial Expressions (DANVA-2-CF)	Child facial expression recognition (total)			1.14 ^b	Significant improvement in identifying child facial expressions.	

Note. All effect sizes were reported to 2 decimal places.

^a Computed using Campbell collaboration effect size calculator (Wilson, n.d.).

^b Computed using Lenhard, W. & Lenhard, A. (2016) from raw data in studies.

^c Effect size interpretation: 0-0.2 (minimal), 0.2- 0.49 (small), 0.5- 0.79 (medium), ≥ 0.8 (large) (Cohen, 1992).

Conclusion and Recommendations

The aim of this literature review was to investigate the effectiveness of ToM based group interventions for improving emotion recognition for children with ASD. A total of 101 peer-reviewed journal articles written in English were generated as a result of the search; six studies were included after the title, abstract and full-text search. The weight of evidence (WoE) framework (Gough, 2007) was used to assess the overall fit of the selected studies for addressing the review question.

Effect sizes were extracted from studies or calculated using data given in the studies. Mixed effects were found among the studies. The studies with a high WoE D rating found effect sizes that range from minimal to medium (Begeer et al., 2015; Soorya et al., 2015); two studies with medium WoE D found large effect sizes on some outcomes alongside minimal effects on other outcomes (Begeer et al., 2011; Stichter et al., 2016). There were only two studies (Solomon et al., 2004; Stichter et al., 2012) that did not find mixed results. However, the small sample size (n=18) meant that the large effect sizes in Solomon et al. (2004) could have been prone to superrealisation bias (Slavin & Smith, 2009). The small sample size in Stichter et al. (2012)'s study also meant that there was insufficient power to find a statistically significant result (Cohen, 1992). More information on the curriculum, lesson plans, teaching materials and test implementation is needed to determine whether or not the teaching-to-the-test effect (Volante, 2004) or the ceiling effect had influenced the outcome of these studies. This meant that improvement in the more complex items of the assessments could have been a result of specific teaching, or the lack of improvement in the more basic items could have been due to the lack of room for improvement.

To conclude, a solid evidence base on the effectiveness of ToM based interventions on improving emotion recognition skills for children with ASD has yet to be established.

The following are recommendations on how future research on the review topic can be conducted. First, none of the studies included in the current review were UK based, and the studies were mostly conducted in clinics. More studies on ToM group interventions have to be carried out in a UK school context, in order to find out how the intervention can be generalised across settings (clinic and school) and implemented by school staff with the support of specialists, such as EPs or Speech and Language therapists. Future studies should also aim at defining and using pre and post-intervention measures that can be easily administered by school staff after some training. The post-intervention effect should also be measured immediately after the intervention period to ensure that the effects of interventions are captured at appropriate times (Gersten et al., 2005). If an alternative intervention group is used as a control, the assessor(s) should ideally be blind to the interventions that the groups are receiving to minimise bias. This element of the assessors being blind to the conditions was not being specifically discussed in the reviewed studies. Future studies should also ensure that the fidelity of the implementation is monitored consistently using different methods and that training and supervision are available to school staff before and throughout the intervention.

Educational Psychologists who wish to implement the ToM group interventions in a UK school context should examine the cultural relevance or appropriateness of the materials, as the reviewed interventions were developed in and for the US and Netherlands. As the effects of the intervention were not consistent, it is not recommended for the interventions to be used solely or in replacement of existing

specialist support in schools in the UK. Educational Psychologists should also keep an open mind and work in partnership with school staff to measure the effects of the ToM group interventions using existing valid and reliable tools while noting qualitative changes, which will potentially be beneficial for developing a ToM group intervention programme that will promote more consistent improvement in emotion recognition skills in the UK school context.

References

- American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR*. Washington, DC: American Psychiatric Association.
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (5th ed.)*. Arlington, VA: American Psychiatric Publishing.
- Barker, C., Pistrang, N., & Elliot, R. (2016). *Research Methods in Clinical Psychology: An Introduction for Students and Practitioners (3rd ed.)*. West Sussex, UK: John Wiley & Sons.
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, 21(1), 37-46.
- Baron-Cohen, S. Jolliffe, T. Mortimore, C. and Robertson, M. (1997). Another advanced test of theory of mind: evidence from very high functioning adults with autism or Asperger Syndrome. *Journal of Child Psychology and Psychiatry*, 38, 813-822.
- Baron-Cohen, S. Wheelwright, S. and Hill, J. (2001). The 'Reading the mind in the eyes' test revised version: A study with normal adults, and adults with Asperger Syndrome or High-Functioning autism. *Journal of Child Psychology and Psychiatry*, 42, 241-252.
- Bajgar, J., Ciarrochi, J., Lane, R., & Deane, F. P. (2005). Development of the levels of emotional awareness scale for children (LEAS-C). *British Journal of Developmental Psychology*, 23, 569–586.

- Bauminger, N. (2007). Brief report: Group social-multimodal intervention for HFASD. *Journal of Autism and Developmental Disorders*, 37(8), 1605–1615.
- Bauminger-Zviely, N., Eden, S., Zancanaro, M., Weiss, P. L., & Gal, E. (2013). Increasing social engagement in children with high-functioning autism spectrum disorder using collaborative technologies in the school environment. *Autism*, 17(3), 317–339.
- Begeer, S., Gevers, C., Clifford, P., Verhoeve, M., Kat, K., Hoddenbach, E., & Boer, F. (2011). ToM training in children with autism: A randomized controlled trial. *Journal of Autism and Developmental Disorders*, 41(8), 997–1006.
- Begeer, S., Howlin, P., Hoddenbach, E., Clauser, C., Lindauer, R., Clifford, P., Gevers, C., Boer, F., & Koot, H. M. (2015). Effects and Moderators of a Short ToM Intervention for Children with Autism Spectrum Disorder: A Randomized Controlled Trial. *Autism Research : Official Journal of the International Society for Autism Research*, 8(6), 738–48.
- Bell, K. S., & Kirby, J. R. (2002). Teaching emotion and belief as mindreading instruction for children with autism. *Developmental Disabilities Bulletin*, 30(1), 16–50.
- Bernad-Ripoll, S. (2007). Using a self-as-model video combined with social stories™ to help a child with Asperger syndrome understand emotions. *Focus on Autism and Other Developmental Disabilities*, 22(2), 100–106.

- Capps, L., Yirmiya, N., & Sigman, M. (1992). Understanding of simple and complex emotions in non-retarded children with autism. *Journal of Child Psychology and Psychiatry*, 33(7), 1169-1182.
- Celani, G., Battacchi, M. W., & Arcidiacono, L. (1999). The understanding of the emotional meaning of facial expressions in people with autism. *Journal of Autism and Developmental Disorders*, 29(1), 57–66.
- Charman, T., & Stone, W. (2008). *Social and Communication Development in Autism Spectrum Disorders: Early Identification, Diagnosis, and Intervention*. New York: Guilford Press.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155.
- Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R.O., Hornik, R. C., Phillips, D. C., Walker, D. F., & Weiner, S. S. (1980). *Toward Reform of Program Evaluation: Aims, Methods, and Institutional Arrangements*. San Francisco: Jossey-Bass.
- Department for Education (DfE) and Department of Health (DH). (2015). *SEND Code of Practice: 0 to 25 Years*. London: The Stationery Office.
- Department for Education (DfE). (2017). *National Statistics: Special Educational Needs in England: January 2017*. Darlington: Crown.
- Feng, H., Lo, Y., Tsai, S., & Cartledge, G. (2008). The effects of theory-of-mind and social skill training on the social competence of a sixth-grade student with autism. *Journal of Positive Behavior Interventions*, 10(4), 228–242.

- Fletcher-Watson, S., McConnell, F., Manola, E., & McConachie, H. (2014). *Interventions Based on the Theory of Mind Cognitive Model for Autism Spectrum Disorder (ASD)*. The Cochrane Library.
- Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71*(2), 149-164.
- Gevers, C., Clifford, P., Mager, M., & Boer, F. (2006). Brief report: a theory-of-mind-based social-cognition training program for school-aged children with pervasive developmental disorders: An open study of its effectiveness. *Journal of Autism and Developmental Disorders, 36*(4), 567–571.
- Gough, D. (2007). Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. *Special Edition of Research Papers in Education, 22*(2), 213-228.
- Gutman, S. A., & Murphy, S. L. (2012). From the desk of the editor and associate editor—information commonly unreported in intervention effectiveness studies. *American Journal of Occupational Therapy, 66*, 7–10.
- Hadwin, J., Baron-Cohen, S., Howlin, P., & Hill, K. (1997). Does teaching theory of mind have an effect on the ability to develop conversation in children with autism? *Journal of Autism and Developmental Disorders, 27*(5), 519–537.

- Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychology Review, 20*(3), 290–322.
- Hoddenbach, E., Koot, H. M., Clifford, P., Gevers, C., Clauser, C., Boer, F., & Begeer, S. (2012). Individual differences in the efficacy of a short theory of mind intervention for children with autism spectrum disorder: a randomized controlled trial. *Trials, 13*, 206.
- Kim, E., Kyeong, S., Cheon, K. A., Park, B., Oh, M. K., Chun, J. W., Park, H. J., Kim, J.J., & Song, D. H. (2016). Neural responses to affective and cognitive theory of mind in children and adolescents with autism spectrum disorder. *Neuroscience Letters, 621*, 117-125.
- Kuusikko, S., Haapsamo, H., Jansson-Verkasalo, E., Hurtig, T., Mattila, M. L., Ebeling, H., Jussila, K., Bolte, S., & Moilanen, I. (2009). Emotion recognition in children and adolescents with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 39*(6), 938-945.
- Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (2003). *Autism Diagnostic Observation Schedule*. Los Angeles, CA: Western Psychological Services.
- Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24*, 659–685.

- Lenhard, W. & Lenhard, A. (2016). *Calculation of Effect Sizes*. Retrieved February 11, 2018, from https://www.psychometrica.de/effect_size.html. Dettelbach (Germany): Psychometrica.
- Nowicki, S., & Carton, J. (1992). The measurement of emotional intensity from facial expressions. *Journal of Social Psychology, 133*, 749–750.
- Peterson, C. C., & Wellman, H. M. (2009). From fancy to reason: Scaling deaf and hearing children's understanding of theory of mind and pretence. *British Journal of Developmental Psychology, 27*(2), 297-310.
- Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: Horses for courses. *Journal of Epidemiology and Community Health, 57*(7), 527–529.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and Brain Sciences, 1*(4), 515-526.
- Rice, L. M., Wall, C. A., Fogel, A., & Shic, F. (2015). Computer-assisted face processing instruction improves emotion recognition, mentalizing, and social skills in students with ASD. *Journal of Autism and Developmental Disorders, 45*(7), 2176-2186.
- Shahaeian, A., Peterson, C. C., Slaughter, V., & Wellman, H. M. (2011). Culture and the sequence of steps in theory of mind development. *Developmental Psychology, 47*(5), 1239.
- Silver, M., & Oakes, P. (2001). Evaluation of a new computer intervention to teach people with autism or Asperger syndrome to recognize and predict emotions in others. *Autism, 5*(3), 299–316.

- Solomon, M., Goodlin-Jones, B. L., & Anders, T. F. (2004). A Social Adjustment Enhancement Intervention for High Functioning Autism, Asperger's Syndrome, and Pervasive Developmental Disorder NOS. *Journal of Autism and Developmental Disorders*, 34(6), 649–668.
- Soorya, L. V., Siper, P. M., Beck, T., Soffes, S., Halpern, D., Gorenstein, M., Kolevzon, A., Buxbaum, J., Wang, A. T. (2015). Randomized comparative trial of a social cognitive skills group for children with autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, 54(3), 208–216.
- Slavin, R., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational Evaluation and Policy Analysis*, 31(4), 500-506.
- Steeneman, P. (1994). [*Learning to Think about Thinking and Learning to Understand Emotions: Group Treatment of Children*] *Leren Denken over Denken en Leren Begrijpen van Emoties: Groepsbehandeling van Kinderen*. Leuven: Garant.
- Steeneman, P., Jackson, S., Pelzer, H., & Muris, P. (1996). Children with social handicaps: An intervention programme using a Theory of Mind approach. *Clinical Child Psychology and Psychiatry*, 1, 251–263.
- Stichter, J. P., Herzog, M. J., Visovsky, K., Schmidt, C., Randolph, J., Schultz, T., & Gage, N. (2010). Social competence intervention for youth with Asperger syndrome and high functioning autism: An initial investigation. *Journal of Autism and Developmental Disorders*, 40, 1067–1079.

- Stichter, J. P., Herzog, M. J., Owens, S. A., & Malugen, E. (2016). Manualization, feasibility, and effectiveness of the school-based Social Competence Intervention for Adolescents (SCI-A). *Psychology in the Schools, 53*(6), 583–600.
- Stichter, J. P., O'Connor, K. V., Herzog, M. J., Lierheimer, K., & McGhee, S. D. (2012). Social competence intervention for elementary students with Aspergers Syndrome and high functioning autism. *Journal of Autism and Developmental Disorders, 42*(3), 354–366.
- Swettenham, J. (2000). Teaching theory of mind to individuals with autism. In S. Baron-Cohen, H. Tager-Flusberg and D. Cohen (Eds.), *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience* (pp. 442-458). Oxford, UK: Oxford University Press.
- Volante, L. (2004). Teaching to the Test: What Every Educator and Policy-Maker Should Know. *Canadian Journal of Educational Administration and Policy, 35*.
- Waugh, C., & Peskin, Joan. (2015). Improving the social skills of children with HFASD: An intervention study. *Journal of Autism and Developmental Disorders, 45*(9), 2961–2980.
- Wellman, H. M., Baron-Cohen, S., Caswell, R., Gomez, J. C., Swettenham, J., Toye, E., & Lagattuta, K. (2002). Thought-bubbles help children with autism acquire an alternative to a theory of mind. *Autism, 6*(4), 343–363.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development, 72*(3), 655-684.

Wellman, H. M., Fang, F., Liu, D., Zhu, L., & Liu, G. (2006). Scaling of theory-of-mind understandings in Chinese children. *Psychological Science, 17*(12), 1075-1081.

Wellman, H. M., Fang, F., & Peterson, C. C. (2011). Sequential progressions in a theory-of-mind scale: Longitudinal perspectives. *Child Development, 82*(3), 780-792.

Williams, B. T., Gray, K. M., & Tonge, B. J. (2012). Teaching emotion recognition skills to young children with autism: A randomised controlled trial of an emotion training programme. *Journal of Child Psychology and Psychiatry, 53*(12), 1268–1276.

Wilson, D. B., Ph.D. (n.d.). *Practical Meta-Analysis Effect Size Calculator* [Online calculator]. Retrieved February 10, 2018, from <https://www.campbellcollaboration.org/this-is-a-web-based-effect-size-calculator/explore/this-is-a-web-based-effect-size-calculator>

Appendix A: Excluded studies

Table A

Articles excluded at the full-text screening stages

Articles	Exclusion Criteria number(s)
Bauminger-Zviely, N., Eden, S., Zancanaro, M., Weiss, P. L., & Gal, E. (2013). Increasing social engagement in children with high-functioning autism spectrum disorder using collaborative technologies in the school environment. <i>Autism, 17</i> (3), 317–339.	7
Bauminger, N. (2007). Brief report: Group social-multimodal intervention for HFASD. <i>Journal of Autism and Developmental Disorders, 37</i> (8), 1605–1615.	8
Bell, K. S., & Kirby, J. R. (2002). Teaching emotion and belief as mindreading instruction for children with autism. <i>Developmental Disabilities Bulletin, 30</i> (1), 16–50.	8
Bergstrom, R., Najdowski, A. C., Alvarado, M., & Tarbox, J. (2016). Teaching children with autism to tell socially appropriate lies. <i>Journal of Applied Behavior Analysis, 49</i> (2), 405–410.	6
Bernad-Ripoll, S. (2007). Using a self-as-model video combined with social stories™ to help a child with Asperger syndrome understand emotions. <i>Focus on Autism and Other Developmental Disabilities, 22</i> (2), 100–106.	8, 9
Belisle, J., Dixon, M. R., Stanley, C. R., Munoz, B., & Daar, J. H. (2016). Teaching foundational perspective-taking skills to children with autism using the PEAK-T curriculum: single-reversal “I–You” deictic frames. <i>Journal of Applied Behavior Analysis, 49</i> (4), 965–969.	6
Corbett, B. A., Swain, D. M., Coke, C., Simon, D., Newsom, C., Houchins-Juarez, N., Jenson, A., Wang, L., & Song, Y. (2014). Improvement in social deficits in Autism spectrum disorders using a theatre-based, peer-mediated intervention. <i>Autism Research, 7</i> (1), 4–16.	6
de Veld, D. M. J., Howlin, P., Hoddenbach, E., Mulder, F., Wolf, I., Koot, H. M., Lindauer, R., & Begeer, S. (2017). Moderating effects of parental characteristics on the effectiveness of a theory of mind training for children with autism: A randomized controlled trial. <i>Journal of Autism and Developmental Disorders, 47</i> (7), 1987–1997.	5
Feng, H., Lo, Y., Tsai, S., & Cartledge, G. (2008). The effects of theory-of-mind and social skill training on the social competence of a sixth-grade student with autism. <i>Journal of Positive Behavior Interventions, 10</i> (4), 228–242.	8
Fisher, N., & Happe, F. (2005). A training study of theory of mind and executive function in children with autistic spectrum disorders. <i>Journal of Autism and Developmental Disorders, 35</i> (6), 757–771.	6

- Gevers, C., Clifford, P., Mager, M., & Boer, F. (2006). Brief Report: A Theory-of-Mind-based Social-Cognition Training Program for School-Aged Children with Pervasive Developmental Disorders: An Open Study of its Effectiveness. *Journal of Autism and Developmental Disorders*, 36(4), 567–571. 7
- Gould, E., Tarbox, J., O’Hora, D., Noone, S., & Bergstrom, R. (2011). Teaching children with autism a basic component skill of perspective-taking. *Behavioral Interventions*, 26(1), 50–66. 6, 9
- Grossman, M., Peskin, J., & Juan, V. S. (2013). Thinking about a reader’s mind: Fostering communicative clarity in the compositions of youth with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 43(10), 2376–2392. 6
- Gupta, B. (2015). ToM in autism: A case study. *Psychological Studies*, 60(3), 339–345. 9
- Hadwin, J., Baron-Cohen, S., Howlin, P., & Hill, K. (1997). Does teaching ToM have an effect on the ability to develop conversation in children with autism? *Journal of Autism and Developmental Disorders*, 27(5), 519–537. 7, 8
- Martins, M. P., & Harris, S. L. (2006). Teaching Children with Autism to Respond to Joint Attention Initiations. *Child & Family Behavior Therapy*, 28(1), 51–68. 6
- Matthews, N. L., Goldberg, W. A., & Lukowski, A. F. (2013). Theory of Mind in children with autism spectrum disorder: Do siblings matter? *Autism Research*, 6(5), 443–453. 5
- McGregor, E., & Bennett, M. (2008). Narrative dependency and the false belief task in autism. *European Journal of Developmental Psychology*, 5(1), 1–18. 6
- McGregor, E., Whiten, A., & Blackburn, P. (1998). 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*, 16(3), 281–300. 6
- Ozonoff, S., & Miller, J. N. (1995). Teaching theory of mind: a new approach to social skills training for individuals with autism. *Journal of Autism and Developmental Disorders*, 25(4), 415–33. 6
- Paynter, J., & Peterson, C. C. (2013). Further evidence of benefits of thought-bubble training for ToM development in children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 7(2), 344–348. 6
- Rice, L. M., Wall, C. A., Fogel, A., & Shic, F. (2015). Computer-assisted face processing instruction improves emotion recognition, mentalizing, and social skills in students with ASD. *Journal of autism and developmental disorders*, 45(7), 2176-2186. 8
- Silver, M., & Oakes, P. (2001). Evaluation of a new computer intervention to teach people with autism or Asperger syndrome to recognize and predict emotions in others. *Autism*, 5(3), 299–316. 8
- Swettenham, J. G., Baron-Cohen, S., Gomez, J.-C., & Walsh, S. (1996). What’s inside someone’s head? Conceiving of the mind 6

- as a camera helps children with autism acquire an alternative to a theory of mind. *Cognitive Neuropsychiatry*, 1(1), 73–88.
- Swettenham, J. (1996). Can children be taught to understand false belief using computers? *Child Psychology & Psychiatry & Allied Disciplines*, 37(2), 157–165. 6
- Waugh, C., & Peskin, Joan. (2015). Improving the social skills of children with HFASD: An intervention study. *Journal of Autism and Developmental Disorders*, 45(9), 2961–2980. 7
- Wellman, H. M., Baron-Cohen, S., Caswell, R., Gomez, J. C., Swettenham, J., Toye, E., & Lagattuta, K. (2002). Thought-bubbles help children with autism acquire an alternative to a theory of mind. *Autism: The International Journal of Research and Practice*, 6(4), 343–63. 6
- Williams, B. T., Gray, K. M., & Tonge, B. J. (2012). Teaching emotion recognition skills to young children with autism: A randomised controlled trial of an emotion training programme. *Journal of Child Psychology and Psychiatry*, 53(12), 1268–1276. 8, 9
-

Appendix B: Weight of Evidence Judgement Criteria and Ratings

Weight of Evidence A (WoE A): Methodological Quality

A high, medium or low WoE A rating is assigned to each study based on the number of essential and desirable criteria that a study has fulfilled (Table B1).

Table B2 outlined the number of essential and desirable criteria fulfilled by each study and the WoE A rating that each study received.

Table B1

Weight of evidence A judgement: Methodological Quality

WoE A Rating	Criteria	Rationale
3 (high)	≥ 9 essential criteria and ≥ 4 desirable criteria	Based on the Gersten et al. (2005)'s coding protocol: judgement criteria for high and acceptable quality research.
2 (medium)	≥ 9 essential criteria and ≥ 1 desirable criteria	
1 (low)	<9 essential criteria and <1 desirable criteria	

Table B2

Weight of Evidence A rating: Methodological Quality

Study	Essential Criteria	Desirable Criteria	WoE A
Begeer et al. (2011)	8	4	1 (low)
Begeer et al. (2015)	9	4	3 (high)
Solomon et al. (2004)	6	5	1 (low)
Soorya et al. (2015)	9	6	3 (high)
Stichter et al. (2012)	5	2	1 (low)
Stichter et al. (2016)	6	2	1 (low)

Weight of Evidence (WoE B): Appropriateness of Design

The WoE B judgement criteria were outlined in Table B3. Systematic review was mentioned to be of the highest methodological relevance but was not suitable for the purpose of the current review, so it was excluded from the judgement table.

WoE B rating for each study is listed in Table B4.

Table B3

Weight of evidence B judgement: Methodological relevance

WoE B Rating	Types of studies	Rationale
3 (high)	Randomised clinical trial studies (random assortment of participants, pre/ post-intervention outcome)	Based on the typology of evidence (Petticrew & Roberts, 2003).
2 (medium)	Cohort studies, Quasi-Experimental design studies	
1 (low)	Case-control studies, Cross-sectional surveys, Qualitative research, Non-experimental evaluation, Case reports	

Table B4

Weight of evidence B rating: Methodological relevance

Study	Design	WoE B
Begeer et al. (2011)	RCT (waitlist control)	3 (high)
Begeer et al. (2015)	RCT (waitlist control)	3 (high)
Solomon et al. (2004)	Quasi-experimental design (individual matching on full-scale IQ and age, then randomly allocated to treatment or waitlist control group)	2 (medium)
Soorya et al. (2015)	RCT (control engaged in facilitated play)	3 (high)
Stichter et al. (2012)	Quasi-experimental design (One group pretest- posttest design)	2 (medium)

Stichter et al. (2016)	Quasi-experimental design (One group pretest- posttest design; longitudinal data grouped as pre-post)	2 (medium)
------------------------	---	------------

Weight of Evidence (WoE C): Topic Relevance (criteria)

The total WoE C was calculated by averaging the scores for A) Intervention setting, B) training and supervision and C) Curriculum (manualised/ standardised).

The studies that had a WoE C rating of 2.33 to 3 will receive a high overall rating, 1.68 to 2.32 will receive a medium overall rating, and 1.67 or below will receive a low overall rating.

Table B5

Weight of evidence C judgement: Topic Relevance

Criteria	Rating	Rationale
A) Intervention setting (ecological validity)	3- School setting 2- Clinic or outpatient clinic setting 1- Not school or Clinic or did not specify	Studies conducted in a more naturalistic environment will have higher ecological validity.
B) Training and supervision provided to intervention administrator	3- Training and supervision 2- Training or supervision only 1- No training or supervision or no information was given	The effectiveness of the interventions can be better explained by the active ingredients/ methods of the interventions, instead of other variables such as provider differences (Gutman & Murphy, 2012).
C) Curriculum relevance: proportion of lessons covering the theme of emotion recognition	3- ≥ 30% of the sessions 2- < 20-30% of the sessions 1- <20% or not specified	This is to investigate the extent to which a curriculum or an intervention focuses on developing emotion recognition skills.

Table B6

Weight of evidence C rating: Topic Relevance

Study	A) Intervention setting	B) Training and supervision of administrator	C) Curriculum relevance (to emotion identification)	WoE C
Begeer et al. (2011).	1 (low)	3 (high)	1 (low)	1.67 (low)
Begeer et al. (2015)	2 (medium)	2 (medium)	2 (medium)	2 (medium)
Solomon et al. (2004)	2 (medium)	2 (medium)	3 (high)	2.33 (high)
Soorya et al. (2015)	1 (low)	3 (high)	2 (medium)	2 (medium)
Stichter et al. (2012)	2 (medium)	1 (low)	3 (high)	2 (medium)
Stichter et al. (2016)	3 (high)	3 (high)	3 (high)	3 (high)

Appendix C: Coding Protocol Example

Quality Indicators for Group Experimental and Quasi-Experimental Research in Special Education (Gersten et al., 2005)

Date: 10th February 2018

Reference: Solomon, M., Goodlin-Jones, B. L., & Anders, T. F. (2004). A Social Adjustment Enhancement Intervention for High Functioning Autism, Asperger's Syndrome, and Pervasive Developmental Disorder NOS. *Journal of Autism and Developmental Disorders*, 34(6), 649–668.

Intervention: Social Adjustment Enhancement Intervention

Design: Quasi-experimental: individually matched on FSIQ and age, and then randomised

Type of Publication: Peer-reviewed journal

Essential Quality Indicators (E)

Quality Indicator for Describing Participants

1. Was sufficient information provided to determine/confirm whether the participants demonstrated the disability(ies) or difficulties presented?

- Yes
 No
 Unable to Code

2. Were appropriate procedures used to increase the likelihood that relevant characteristics of participants in the sample were comparable across conditions?

- Yes
 No
 Unable to Code

3. Was sufficient information given characterising the interventionists or teachers provided? Did it indicate whether they were comparable across conditions?

- Yes
 No
 Unable to Code

Implementation of the Intervention and Description of Comparison Conditions

4. Was the intervention clearly described and specified?

- Yes
 No

Unable to Code

5. Was the fidelity of implementation described and assessed?

- Yes
 No
 Unable to Code

6. Was the nature of services provided in comparison conditions described?

- Yes
 No
 Unable to Code

Quality Indicators for Outcome Measures

7. Were multiple measures used to provide an appropriate balance between measures closely aligned with the intervention and measures of generalised performance?

- Yes
 No
 Unable to Code

8. Were outcomes for capturing the intervention's effect measured at the appropriate times?

- Yes
 No
 Unable to Code

Quality Indicators for Data Analysis

9. Were the data analysis techniques appropriately linked to key research questions and hypotheses? Were they appropriately linked to the unit of analysis in the study?

- Yes
 No
 Unable to Code

10. Did the research report include not only inferential statistics but also effect size calculations?

- Yes

No

Unable to Code

Desirable Quality Indicators (D)

1. Was data available on attrition rates among intervention samples? Was severe overall attrition documented? If so, is attrition comparable across samples? Is overall attrition less than 30%?

- Yes
- No
- Unable to Code

2. Did the study provide not only internal consistency reliability but also test-retest reliability and interrater reliability (when appropriate) for outcome measures? Were data collectors and/or scorers blind to study conditions and equally (un)familiar to examinees across study conditions?

- Yes
- No
- Unable to Code

3. Were outcomes for capturing the intervention's effect measured beyond an immediate post-test?

- Yes
- No
- Unable to Code

4. Was evidence of the criterion-related validity and construct validity of the measures provided?

- Yes
- No
- Unable to Cod

5. Did the research team assess not only surface features of fidelity implementation (e.g. number of minutes allocated to the intervention or teacher/interventionist following procedures specified), but also examine the quality of implementation?

- Yes
- No
- Unable to Code

6. Was any documentation of the nature of instruction or series provided in comparison conditions?

- Yes
- No
- Unable to Code

7. Did the research report include actual audio or videotape excerpts that capture the nature of the intervention?

- Yes
- No
- Unable to Code

8. Were results presented in a clear, coherent fashion?

- Yes
- No
- Unable to Code

Overall score: E6 D5

WoE A rating: 1 (low)

