Case Study 1: An Evidence-Based Practice Review Report
Theme: School (setting) based interventions for children with special educational needs (SEN)

The effectiveness of Mind Reading: An Interactive Guide to Emotions in increasing the emotion recognition ability of children aged 6 to 12 with Autism Spectrum Conditions

Summary

This systematic literature review and meta-analysis examined the effectiveness of Mind Reading: An Interactive Guide to Emotions (Baron-Cohen, 2003) in increasing the emotion recognition ability of children with Autism Spectrum Conditions (ASC). Mind Reading is an educational interactive computer-based guide to teaching a broad range of basic and complex emotions in the face and voice. Children engage with the programme individually. The programme consists of three main components: an Emotions Library, a Learning Centre, and a Games Zone.

A comprehensive literature search found ten studies that met the inclusion criteria of this review. The ten studies were then reviewed using Gough's (2007) Weight of Evidence (WoE) Framework. The findings from meta analysing all ten studies suggest that Mind Reading: An Interactive Guide to Emotions can significantly increase facial emotion recognition in children with ASC, with medium to large effect sizes reported. The review addresses both strengths and limitations of existing evidence.

Introduction

Mind Reading: An Interactive Guide to Emotions, available on DVD-ROM and CD-ROM, is an interactive computer-based guide to reading a broad range of basic and complex emotions from the face and voice. Mind Reading is an educational programme designed to teach emotion recognition and understanding of others

through videos of facial expressions and tone of voice. The programme was not designed to target a specific audience but young people with ASC have become the natural audience. One CD/DVD can be purchased for between £20.00 - £29.99.The programme does not allow multiple users simultaneously but has an unlimited number of uses. Mind Reading is comprised of 412 emotions, organized into 24 emotion groups and by 6 emotion levels which correspond with ages. For example level one is suggested for ages 4-7 and level 6 is suggested for use with individuals over the age of 18. These groups contain common emotions but also ‘sub-emotions’ which are considered subtle and more difficult to discern.

There are several areas within the programme that provide instruction and reinforcement. These areas include the Emotions Library, Learning Centre, Games Zone, and Rewards Zone. In the Emotions Library, children chose an emotion group and observe six different actors who represent genders male or female, different ages and different cultures. The video clip of the facial expression is accompanied by a voiceover describing the facial characteristics and emotions are also defined in text vignettes. The Learning Centre contains lessons to teach emotion recognition with video and audio examples. Prior to, and after each of the lessons, there are quizzes. Through the quizzes, rewards can be earned and collected. The Game Zone is designed to allow users to practice the targeted emotion recognition skills. The activities focus on Hidden Face, Space Face, Emotion Pairs, Real World Face and Famous Face games. In the Famous Face games through the use of the intensity slider, the user has active control of the emotion displayed and can ‘make’ an actor display an emotion, for example, happy or sad. Users can alter the emotions of the actor Daniel Radcliffe (who plays Harry Potter). In Real World face, scenes from

offices, schools or markets are shown and the user drags speech bubbles to the actor which changes their corresponding emotional state.

The Mind Reading programme incorporates an internal reinforcement system where users accrue “rewards” for successfully completing tasks and quizzes in the learning zone and game zone. The “rewards” are then used to unlock pictures (such as musical instruments or flags of the world) in the Rewards Zone. When the pictures are clicked in the Reward Zone users can see videos and additional information. There is a range of rewards that may capitalise on the special interests of children with ASC to encourage them to use the programme often.

Mind Reading has a corresponding handbook which gives comprehensive details of the programme for the facilitator and details of online help. Within the programme, there is a function called the MindReading Manager which allows the teacher/facilitator to configure the program, such as setting time limits in games and limiting emotions (for example removing the ‘romantic’ emotions for younger children). This omnipresent control also allows the teacher/parent/facilitator to monitor progress such as overall engagement with different components, lists of emotions completed and average scores.

Psychological Basis

It is generally accepted that there are six basic expressions of facial affect; happiness, surprise, fear, anger, sadness and disgust (Ekman & Friesen, 1976). It has been suggested that the ability to recognise emotions and understand mental states begins to emerge in the first year of life (Heck, et al., 2018). At around two

years old children can understand intentions (Sodian & Thoermer, 2008) and by age three typically developing children can accurately identify and label these six emotions (Iordanou & Mattock, 2021; Widen & Russell, 2003). The development of emotion recognition skills continues throughout childhood and adolescence and is an imperative component in the development of more complex social perception such as mentalising ability and Theory of Mind (TOM; Ashwin, et al.,2006) often referred to as 'mind reading' or 'empathising' (Baron-Cohen, 2002; 2003). Children with ASC have been shown to display less accuracy in their ability to identify facial expressions that convey emotion than typically developing children (Harms, et al., 2010; Paynter & Peterson, 2010; Schultz, 2005). These findings have been reported for both children with high-functioning ASC (Bal et al., 2010) and children with ASC who have a lower range of cognitive ability (Gross, 2004; Tardif, et al., 2007).

The use of computer software to deliver educational programmes for individuals with ASC is said to have a number of advantages such as individuals with ASC preferring the consistent and predictable nature of computer software which is also free from social demands. Working individually on a computer means the user can work at their own pace and that lessons can be repeated innumerable times until satisfied with outcomes (Bishop, 2003; Moore, et al., 2000)

In the Mind Reading programme, learning is based on repetition and retrieval. It is well accepted that memory performance is enhanced and maintained after repetition learning (for example Ebbinghaus', 1964 'learning effect'). Karpicke and Roediger, (2008) demonstrated the critical role of retrieval in learning as it enhanced long term retention. Throughout each area in the Mind Reading programme users have an

opportunity to repeat and consolidate their learning as well as execute retrieval practice in the Gaming Zone. However, it may be important to consider if transferability of the emotion recognition skills achieved through the Mind Reading programme is possible. Especially as Karpicke et al. (2016) have also proposed that retrieval practice effects can be best explained as the “episodic context account” which posits that retrieval practice prompts individuals to reinstate the context in which information was acquired (Whiffen & Karpicke, 2017). The context of every day life may not be close enough to the context the child experiences engaging in the Mind Reading intervention.

Rationale and Relevance

In the UK ASC remains a prevalent primary classification of special educational need among students with Education, Health and Care plans or statements (Department for Education, 2019). A core characteristic of ASC is social deficits (American Psychiatric Association, 2013). Reduced capacity to recognise and respond to others’ emotional expressions is said to precipitate these difficulties (Baron-Cohen, et al., 2009). Moreover, experiencing difficulty in facial emotion recognition has also been linked to frequent displays of social and interpersonal behaviour that is considered challenging or antisocial (Marsh & Blair, 2008). Therefore developing and enhancing emotion recognition skills of children and young people with ASC is a common and critical consideration for educational psychologists. Recent guidance from the Education Endowment Foundation (Wigelsworth et al., 2020) regarding social and emotional learning for all children advocates strategies to develop children’s social awareness through teaching strategies to recognise emotions.

Social exposure and interaction account for a substantial amount of implicit learning about emotion recognition thus the impact of the Covid-19 pandemic must be acknowledged. Since the onset of the pandemic, all children experienced a reduction in the number of opportunities to engage in social interactions. In addition to this, during current social interactions, there is an enduring reduction of accessibility to facial information (due to social distancing and mask-wearing). As Mind Reading is a computer-based programme in which young people can engage individually, without the constant or close presence of a facilitator it presents a potentially pertinent solution to remediation of emotion recognition deficits in ASC, especially in the current context. A cost-effective technology-based intervention devised to improve emotion recognition skills in ASC is therefore of particular interest to educational psychologists. In Fonagy et al.'s (2014) seminal work 'What Works for Whom,' Mind Reading is noted as a 'promising' intervention (p. 297) with further research sought after. Accordingly, the current review and meta-analysis explores the effectiveness of Mind Reading: An Interactive Guide to Emotions in improving facial emotion recognition in children with ASC.

Review question

How effective is Mind Reading: The Interactive Guide to Emotions 1.3 in improving facial emotion recognition skills for children aged 6 to 12 with Autism Spectrum Conditions?

Critical Review of the Evidence Base

Literature search

On the 16th of December 2021, a literature search was conducted using three online electronic databases: Eric (EBSCO), Web of Science and PsycINFO. The search terms are shown in table 1.

Table 1

Search Terms

autis* OR Asperger* OR ASD* OR ASC* OR HFASD OR LFASD OR "autism spectrum disorder*" OR "autism spectrum condition*" OR "pervasive development disorder*" OR PDD	OR	emotion*" OR "face*" or "facial*" OR feeling* OR "expression*" OR "emotion* recog*" OR "recognition of feel	AND	Mindread* OR Computer* OR software* OR technolog* OR "computer based" OR "computer assisted" OR "computer game*" OR intervention OR training

The initial search identified 1455 papers and imported them into a database (OneNote). Due to duplication, 1353 papers were excluded; all papers were then screened according to the exclusion criteria (see table 2). This led to the exclusion of 86 papers with 16 papers then screened for titles and abstracts. A final full-text

screening of ten studies was conducted with these ten studies found to be eligible for inclusion in the study (see Figure 1 for a flow diagram of the screening process; see Table 3 for final study selection).

Table 2

Inclusion and Exclusion Criteria

Study Feature	Inclusion	Exclusion	Rationale
1. Type of publication	The study must be published in a peer reviewed journal.	Grey literature such as dissertations were excluded.	This will ensure high standard research is included in the review which has undergone rigorous scrutiny
2. Language of publication	English	Studies published in other languages	The reviewer is only fluent in English.
3. Type of study	Experimental in design, including randomised controlled trials, quasi experimental design and pre and post-test design and	Studies which do not implement experimental conditions or are qualitative	The reviewer is only interested in quantitative studies. This will allow a measure of the

single case experimental designs	or correlational. Reviews, articles and meta-analyses are also excluded. Follow up studies are also excluded if they include the same sample or data set as studies already selected for the review.	effectiveness of the interventions and enable comparisons between studies using pre intervention and post intervention data. Examining studies which comprise experimental designs will enable examination of the effectiveness of Mind Reading using between or within participant comparisons.

4. Participants	Participants diagnosed with ASD under the Diagnostic Statistical Manual of Mental Disorders 5 (DSM-5; 2013) or IV (DSM-IV; 1994), International Classification of Diseases 10 (ICD-10) (World Health Organization, 1992) or the older versions. Studies targeting participants with Autistic Disorder, Asperger's Syndrome and Pervasive Developmental Disorder Not Otherwise Specified were included. Ages Under 18 years old	Individuals with Rett syndrome as the primary focus were excluded as the symptoms associated with this condition are related to a specific gene mutation. Children who do not have a diagnosis of Autistic Disorder, Asperger's Syndrome and Pervasive Developmental Disorder Not Otherwise Specified under DSM-5, DSM-	This review focuses on the effectiveness of Mind Reading for remediating difficulties in facial emotion recognition in children with Autism Spectrum Conditions, Asperger's or Pervasive Developmental Disorder

		IV (American Psychiatric Association, 2013), ICD-10 (World Health Organization, 1992) or the older versions	
		Adults and individuals over 18 years old	
5. Type of intervention	Intervention named: ‘Mind Reading: The Interactive Guide to Emotions 1.3’ by Baron-Cohen (2003) available on CD-ROM/DVD-ROM	Studies that do not include the intervention: ‘Mind Reading: The Interactive Guide to Emotions 1.3’ by Baron-Cohen (2003)	‘Mind Reading: The Interactive Guide to Emotions 1.3’ by Baron-Cohen (2003) is the intervention of interest

Figure 1 Flow Diagram of Selection and Screening Process

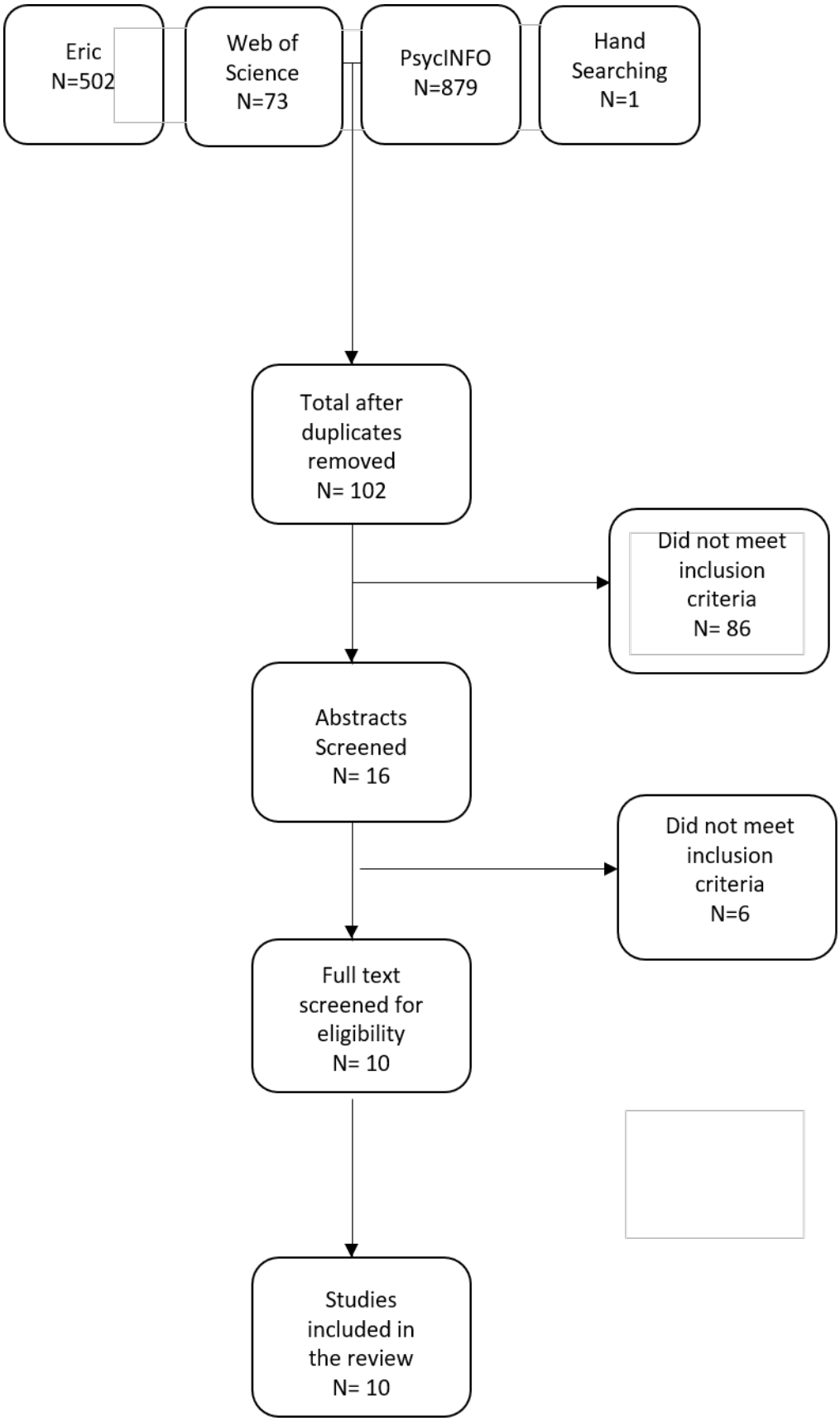


Table 3

The Final Ten Studies Included in the Systematic Review

Davidson, D., Hilvert, E., Winning, A. M., & Giordano, M. (2021). Recognition of Emotions from Situational Contexts and the Impact of a Mind Reading Intervention in Children with Autism Spectrum Disorder. *Child Psychiatry and Human Development*. https://doi.org/10.1007/s10578-021-01139-0

Lacava, P., Golan, O., Baron-Cohen, S., & Smith Myles, B. (2007). *Using Assistive Technology to Teach Emotion Recognition to Students With Asperger Syndrome*. 174–181. https://doi.org/10.1177/07419325070280030601

Lacava, P. G., Rankin, A., Mahlios, E., Cook, K., & Simpson, R. L. (2010). A single case design evaluation of a software and tutor intervention addressing emotion recognition and social interaction in four boys with ASD. *Autism*, *14*(3), 161–178. https://doi.org/10.1177/1362361310362085

Lopata, C., Thomeer, M. L., Volker, M. A., Lee, G. K., Smith, T. H., Smith, R. A., Mcdonald, C. A., Rodgers, J. D., Lipinski, A. M., & Toomey, J. A. (2012). Feasibility and Initial Efficacy of a Comprehensive School-Based Intervention for High-Functioning Autism Spectrum Disorders. *Psychology in the Schools*, *49*(10), 963–974. https://doi.org/10.1002/pits.21649

Lopata, C., Thomeer, M. L., Rodgers, J. D., Donnelly, J. P., & McDonald, C. A. (2016). RCT of mind reading as a component of a psychosocial treatment for high-

functioning children with ASD. *Research in Autism Spectrum Disorders*, *21*, 25–36. https://doi.org/10.1016/j.rasd.2015.09.003

Lopata, C., Thomeer, M. L., Rodgers, J. D., Donnelly, J. P., McDonald, C. A., Volker, M. A., Smith, T. H., & Wang, H. (2019). Cluster Randomized Trial of a School Intervention for Children with Autism Spectrum Disorder. *Journal of Clinical Child and Adolescent Psychology*, *48*(6), 922–933. https://doi.org/10.1080/15374416.2018.1520121

Lopata, C., Thomeer, M. L., Volker, M. A., Lee, G. K., Smith, T. H., Rodgers, J. D., Smith, R. A., Gullo, G., McDonald, C. A., Mirwis, J., & Toomey, J. A. (2013). Open-Trial Pilot Study of a Comprehensive School-Based Intervention for High-Functioning Autism Spectrum Disorders. *Remedial and Special Education*, *34*(5), 269–281. https://doi.org/10.1177/0741932512450518

Thomeer, M., Rodgers, J., Lopata, C., McDonald, C., Volker, M., & Toomey, J. et al. (2011). Open-Trial Pilot of Mind Reading and In Vivo Rehearsal for Children With HFASD. Focus On Autism And Other Developmental Disabilities, 26(3), 153-161. doi: 10.1177/1088357611414876

Thomeer, M. L., Smith, R. A., Lopata, C., Volker, M. A., Lipinski, A. M., Rodgers, J. D., McDonald, C. A., & Lee, G. K. (2015). Randomized Controlled Trial of Mind Reading and In Vivo Rehearsal for High-Functioning Children with ASD. *Journal of Autism and Developmental Disorders*, *45*(7), 2115–2127. https://doi.org/10.1007/s10803-015-2374-0

Weinger, P. M., & Depue, R. A. (2011). Remediation of deficits in recognition of facial emotions in children with Autism spectrum disorders. *Child and Family Behavior Therapy*, *33*(1), 20–31. https://doi.org/10.1080/07317107.2011.545008

Weight of Evidence

The final ten studies were evaluated according to the WoE framework (Harden & Gough, 2012) which outlines three dimensions to judge the studies against:

WoE A (WoE A) which evaluates methodological soundness; the coding protocol from Gersten et al.'s (2005) was used to assess Randomised Control Trials (RCTs), group experimental designs and the quasi-experimental design study. Horner et al.'s (2005) framework was used to review the single-case experimental design studies. Full details can be found in appendices A to D.

WoE B (WoE B) considers methodological relevance in relation to the review question using the hierarchy from Petticrew and Roberts (2003).

WoE C (WoE C) assesses the relevance of the study to the review question.

Scores for each dimension were then averaged to produce a final overall WoE D (WoE D – see Table 4).

Table 4

Overall Weight of Evidence Ratings

Authors	WoE A	WoE B	WoE C	WoE D	WoE D descriptors
Davidson et al., 2021 Study 2	1	2	2	1.67	Low
Lopata et al., 2013	1	2	3	2	Medium
Lopata et al., 2019	3	3	2	2.67	High
Lopata et al., 2016	3	3	2	2.67	High
Lopata et al., 2012	3	2	2	2.34	Medium
Thomeer et al., 2011	3	2	1	2	Medium
Thomeer et al., 2015	1	3	3	2.34	Medium
Weinger & Depue, 2011	1	2	2	1.67	Low

Lacava et al., 2007	1	2	3	2	Medium
Lacava et al., 2010	1.7	2	3	2.34	Medium

WoE A score descriptor: High 3, Medium 2, Low 1
WoE B score descriptor: High 3, Medium 2, Low 1
WoE C score descriptor: High score 7 to 10 equates to 3, Medium score 4 to 6 equates to 2, Low score 1 to 3 equates to 1
WoE D score descriptor: High greater than 2.5, Medium 2 to 2.5, Low less than 2

Critical Review of Included Studies

Participants

Participants were all enrolled in various school settings (N = 265). Their ages ranged from 6 – 12 years. A power analysis was carried out in one study (Lopata et al., 2019). This study used a smaller effect size (d = .99) at the 95% confidence interval (CI) [.39, 1.6] which was derived from two pilot studies that had been completed previously and included in this review (Lopata et al., 2012, 2013). To protect statistical power Lopata et al. (2019) projected at 5% dropout rate and adjusted their targeted sample so that they recruited 48 participants per condition.

The inclusion of RCTs, group-based studies and a single case design study meant that there was a large variation in sample size ranging from 4 participants in the single case design (Lacava,et al., 2010) to a Cluster Randomized Trial including 103 participants. Mind Reading was administered in schools in two of the studies, on university campuses in seven studies and at home in one study. All but two studies report participant ethnicity demographics (Weinger & Depue, 2011; Lacava, et al., 2010). None of the studies reported the socioeconomic status of the participant's families. However, six of the studies (Lopata et al., 2012; Lopata et al., 2013; Lopata et al., 2016; Lopata et al., 2019; Thomeer et al., 2015; Thomeer et al., 2011) reported parental education level. Gender ratios were described in each of the

studies. All of the studies included substantially more male participants than females and Lacava, et al's (2010) study included male participants exclusively.

All of the studies were conducted in the United States of America which has a different educational context and curriculum to the United Kingdom (UK). This impedes the adoption of this intervention as 'evidence based' in a UK context. However, eligibility criteria for all of the studies included in the review included a diagnosis of ASC or Pervasive Development Disorder (PDD). In each of the studies, the diagnosis was confirmed using assessments and measures which are routinely administered as part of neurodevelopmental assessments in the UK, for example the Autism Diagnostic Interview – Revised (ADI-R) or Wechsler Intelligence Scale for Children Fourth UK Edition (WISC IV),. Ostensibly, even though the same measures are used to diagnose ASC/PDD in both the UK and USA experiences of children in both cultures are different. However, it may be useful to consider investigating the use of Mind Reading in a UK population. Lopata, et al. (2016) and Thomeer et al. (2015) noted that they recruited participants using public announcements however did not state from which settings. Thomeer et al. (2011) emailed flyers to mental health professionals, local family support groups and directors of special education at schools to recruit participants. Lacava et al. (2007) did not document sampling or recruitment procedures.

Study Design

Five of the ten studies used a pre-test/post-test group experiment design (Davidson et al., 2021 study 2; Lopata et al., 2013; Lopata et al., 2012; Thomeer et al., 2011; Lacava et al., 2007) without control groups resulting in lower WoE B scores for all

due to difficulties in reliably extricating effects of Mind Reading from other confounding factors such as the effect of engaging in school, or other agencies in studies with more than one school or service. One study had a Single Case design (Lacava et al., 2010). A further two studies used an RCT (Lopata et al., 2016; Thomeer et al., 2015), and one study used a Cluster Randomized Trial design (Lopata et al., 2019) which resulted in high ratings for WoE B as the study design is best placed to answer questions about effectiveness due to the high internal validity (Petticrew & Roberts, 2006). Finally, Weinger and Depue's (2011) study comprised a quasi-experimental design however it contained uneven samples and participants in each group were not matched. Only one group received the intervention. Thomeer et al. (2015) used post intervention measures and a 5 week follow up measure to ascertain whether the effects of engaging in Mind Reading were maintained.

Measures

Eight of the studies used the Cambridge Mindreading Face-Voice Battery for Children (CAM-C; Golan, et, al., 2015). The CAM-C measures emotion recognition for 15 emotion concepts using facial expression videos. Six studies (Lopata et al., 2013; Lopata et al., 2019; Lopata, et al., 2016; Lopata et al., 2012; Thomeer et al., 2011; Thomeer et al., 2015) used the Social Responsiveness Scale (SRS; Constantino & Gruber, 2012) which was completed by parents and teachers. The measure has 5 subscales which measure Social Awareness, Social Cognition, Social Communication, Social Motivation, and Autistic Mannerisms. Lopata et al. (2013), Lopata et al. (2016) and Lopata et al. (2012) employed measures of broader social performance completed by both parents and teachers; the Social Skills subscale of the Behaviour Assessment System for Children, Second Edition, (BASC-

2). These studies also used the Adapted Skillstreaming Checklist (ASC; Lopata et al, 2008). Lopata et al. (2016) used a child, parent, teacher and clinician measure. Davidson et al. (2021) and Weinger and Depue, (2011) measured the number of correctly identified emotions before and after engaging in Mind Reading which resulted in lower WoE A and C ratings. Thomeer et al. (2011) did not use the CAM-C but used the Emotion Recognition and Display Survey (ERDS) to determine the number of correctly identified emotions pre and post intervention.

Outcomes and Meta-Analysis

Studies that reported effect sizes did so in terms of Cohen's *d* and ω^2. Weinger and Depue (2011), Lacava et al. (2007) and Lacava et al. (2010) did not report effect sizes however reported sufficient information for effect sizes to be calculated. Studies by Lopata et al. (2019), Lopata et al. (2016) and Thomeer et al. (2015) comprised between-group designs which included an 'intervention' group (who engaged in Mind Reading) and a control group. Lopata et al. (2019) reported a within-group effect size for the intervention group. Weinger and Depue, (2011) did include a control group however the control group and treatment group were not matched on participant characteristics. The intervention group comprised six children with ASC and the control group included 11 children without ASC. Weinger and Depue, (2011) reported only within-group data (pre-test and post-test) about children with ASC who engaged in Mind Reading thus an effect size was calculated for the intervention group. Lopata et al. (2016) and Thomeer et al. (2015) also reported sufficient within-group data about the intervention group to compute effect sizes (in terms of Cohen's d), standard error and confidence intervals to synthesise these studies alongside the other three within-group studies (see table 5 for reported and

calculated effect sizes). In accordance with Cohen (1988), effect sizes in this review are interpreted as small (0.2), medium (0.5) and large (0.8). Large effect sizes were found in all but Lopata et al's (2016) study (d= 0.7). It is important to be aware that increased power is associated with within subject designs which could lead to an overestimation of the true effect size (Dunlap, et al., 1996).

Table 5

Calculated and Reported Effect Sizes

Authors	Number of participants	Measure	Source of effect size	Effect size reported in study	Effect size Cohen's d	Descriptor of Cohen's d	Overall WoE Rating
Davidson et al. (2021; Study 2)	24	Number of correctly identified emotions	Reported and converted using psychometrica	$\eta_p^2 = 0.22$	1.06	Large	Low
Lopata et al. (2013)	12	CAM-C	Reported	d = 0.94	0.94	Large	Medium
Lopata et al. (2019)	103 (52 treatment, 51 control)	CAM-C	Reported	d = 1.41	1.41	Large	High
Lopata et al. (2016)	36 (18 treatment, 18 control)	CAM-C	Reported and Reported and converted using psychometrica	Between-group (intervention and control) effect size $\omega^2 = 0.298$	Calculated from pre and post-intervention group data 0.70	Medium	High
Lopata et al. (2012)	12	CAM-C	Reported	d=1.64	1.64	Large	Medium
Thomeer et al. (2011)	11	ERDS	Reported	d=0.95	0.95	Large	Medium
Thomeer et al. (2015)	43 (22 treatment, 19 control)	CAM-C	Reported and converted using psychometrica	Between-group (intervention and control) effect size $\omega^2 = 0.23$	Calculated from pre and post-intervention group data 1.28	Large	Medium

Weinger & Depue, (2011)	6	Number of correctly identified emotions	Calculated using Campbell Collaboration calculator	Did not report	6.55	Large	Low
Lacava, et al. (2007)	8	CAM-C	Calculated using Campbell Collaboration calculator	Did not report	0.88	Large	Medium
Lacava, et al.(2010)	4	CAM-C	Calculated using Campbell Collaboration calculator	Did not report	2.46	Large	Medium

Note:
All effect sizes above are within group effect sizes. Lopata et al. (2013), Lopata et al. (2019), Lopata et al. (2012) and Thomeer et al. (2011) reported effect sizes in terms of Cohen's d. Davidson et al. (2021; Study 2) reported an effect size in ηp2 which was converted to Cohen's d using psychometrica. Although Lopata et al's (2016) study was an RCT which reported between group effects, Lopata et al. (2016) also reported a within group effects size for the treatment group that engaged in the Mind Reading intervention (size ω^2= 0.298) using pre-intervention scores on the CAM-C and post intervention scores on the CAM-C from the intervention group. This effect size was then converted using psychometrica. Similarly Thomeer et al's (2015) study was a randomized control trial which reported between group effects, Thomeer et al. (2015) also reported a within group effects size for the treatment group that engaged in the Mind Reading intervention (size ω^2= 0.23) using pre intervention scores on the CAM-C and post intervention scores on the CAM-C from the intervention group. This effect size was then converted using psychometrica. Weinger and Depue, (2011), Lacava, et al. (2007) and Lacava, et al., (2010) did not report effect sizes but reported the required information to calculate effect sizes, these were computed using Campbell Collaboration calculator.

Meta-Analysis

A meta-analysis of all but one (Lacava et al., 2010) of the studies in this review was conducted using the meta-essentials package (Suurmond, et al., 2017). It was executed by applying a random-effects method. As can be seen from the forest plot

in figure 2 the meta-analysis revealed a significant and large treatment effect for Mind Reading on Emotion Identification (d=1.07, 95%CI [1.00,1.13]). Results revealed no significant heterogeneity in effect size across all eight studies Q=3.52, I^2=0.0%. (Higgins & Thompson, 2002). Both Q and I^2 are considered low and insignificant scores respectively suggesting studies within the analysis can be considered studies of the same population. Due to publication bias, often studies pursue effects which may mean effect sizes from meta-analyses can be overestimated. To account for this, publication bias is calculated through funnel plots using a random-effects model. If there is no presence of publication bias studies should be distributed normally around the mean effect size. Visual inspection of the funnel plot in this meta-analysis shown in figure 3 revealed asymmetry in the distribution of effect sizes around the mean effect size. Therefore, a further analysis was conducted using Egger's regression test which was not significant (p=0.19). This can be taken to suggest there is no publication bias. It must be noted that the effect size alongside the confidence intervals for the intervention group in Weinger and Depue's (2011) study is untenably large and the standard error in Davidson et al. (2021) study 2 is an exceptionally small value which has influenced the outcome of the meta-analysis and potentially skewed the results, therefore, appendix E contains variations of this meta-analysis which omit each of these studies in turn. Moreover, although Lacava et al. (2010) produced a single case design evaluation of Mind Reading in four boys it was possible to compute an overall group effect size which is shown in the forest plot in appendix F.

Among the 10 intervention studies, 20% (*N* = 2) delivered Mind Reading in school; 10% (*N* = 1) delivered Mind Reading at home and the remaining 70% (*N* = 7) delivered Mind Reading in a research setting. Study design and outcome measures

also varied across studies, with details presented in the mapping table in Appendix A. 70% of studies employed CAM-C as an outcome measure, 10% used ERDS (Thomeer et al., 2011) and 10% used the number of correctly identified emotions (Davidson et al., 2021 study 2) as an outcome measure. Although publication biases are said to always exist in any meta-analysis (Lipsey & Wilson, 1993) the calculation of the publication bias in this meta-analysis through the use of Egger's regression was not significant (p=0.19). This can be taken to suggest there is no publication bias in this instance. Thus, the large effect size of Mind Reading revealed through the use of the random effects analysis is considered to be valid.

Overall this meta-analysis revealed that Mind Reading: An Interactive Guide To Emotions could significantly foster young people with ASC's facial emotion recognition skills. The effect size obtained in the current meta-analysis is large (d=1.07, 95%CI [1.00,1.13]). However, it must be noted that careful interpretation of the findings of the meta-analysis is required as only 4 of the 10 studies (Lopata et al., 2019; Lopata et al., 2016; Lopata et al., 2012; Thomeer et al., 2011) were rated high on WoE A for methodological quality. In addition, only two of the 4 that were rated high on WoE A (Lopata et al., 2019; Lopata et al., 2016) were rated high overall for WoE D (see table 4).

Figure 2

Meta-Analysis including all within-group findings

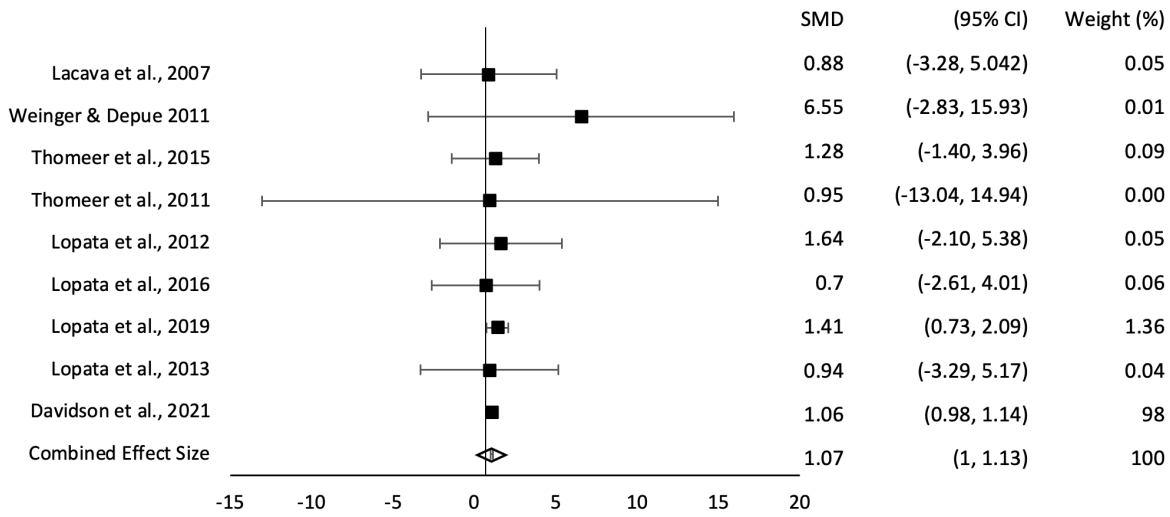
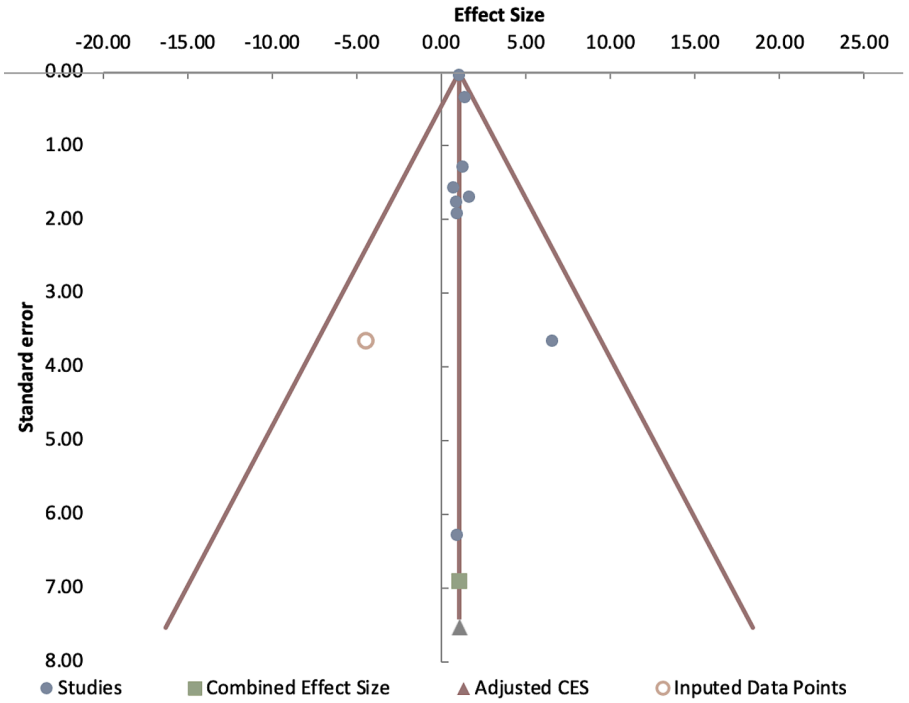


Figure 3

Funnel Plot to Visually Inspect Publication Bias



Conclusion

The purpose of this review was to evaluate the effectiveness of Mind Reading in improving children with ASC's facial emotion recognition. The studies that were

included in this review do show that Mind Reading has a beneficial effect. This is demonstrated by the mainly large effect sizes (Cohen's d) reported by each of the studies in the review. In addition, the meta-analysis result reported in this review was significant. Eight out of the ten studies included in this review were not RCTs, one of these eight studies was a cluster randomized trial (Lopata et al., 2019) and,Lacava et al., (2010) was the only single case design. The other six studies that were not RCTs were quasi-experimental in design (Davidson et al., 2021, Study 2; Lopata et al., 2013; Lopata et al., 2012; Thomeer et al., 2011; Weinger & Depue, 2011; Lacava et al., 2007) Two of the RCTs ran concurrent social skills interventions (the other interventions were inVivo rehearsal and a programme called summerMax). One-group pre-test/post-test designs have inherent validity concerns which mean interpretation of their findings must be considered with caution (Barker, et al., 2016). The review suggests that Mind Reading has promising results for use with children with ASC. However, further experimental research which includes comparison or control groups is required to examine in detail the specific areas of Mind Reading that are likely to account for most of the improved outcomes. In addition to this, including a comparison/control group in future research could elucidate potential confounding variables (such as the effect of school). Although the Mind Reading programme was developed in the UK, all of the studies included in this review were executed in the United States. Ostensibly, to have educational significance in the UK context further studies must be conducted in the UK.

The review provides evidence that indicates Mind Reading is an effective intervention to reduce wider social difficulties experienced by children with ASC. Six of the studies in this review used the SRS(Constantino & Gruber, 2012) completed by parents, teachers or both. The SRS identifies and quantifies the symptoms of

social impairments associated with ASC. A significant reduction in SRS scores following the use of Mind Reading was reported by Lopata et al. (2013), Lopata et al. (2019), Thomeer et al. (2011) and Thomeer et al. (2015). Although a reduction in parents and teacher scores was reported in Lopata et al. (2012), it was not significant for the teacher report. A reduction in the SRS score post-test was reported by Lopata et al. (2019) which was not significant however a significant reduction was reported at follow up which may suggest sustained cumulative effects of Mind Reading on the reduction of social impairments in children with ASC. Indeed, this would require further investigation.

Limitations

Sources of bias may have influenced the results of the studies included in this review; the authors Lopata, Thomeer, Rodgers and McDonald contributed to six of the studies (and Volker contributed to five of those studies, Smith to three, Lee and Lipinski to two). These authors declare no conflict of interest in their studies however it would seem that they have invested a lot of time in the intervention which may make them indifferent to potential flaws of the intervention. Another limitation of the studies included in this review was that all of the studies were conducted in the USA which has a different cultural, medical and educational context to the UK. It has also been said that studies executed in the USA are more difficult to generalise (Nind et al., 2005). A further limitation of this review is all of the studies included, recruited far fewer female participants than males. It must be noted that autism as a diagnosis is more prevalent in males than in females across age groups (Fombonne 2009; Russell et al. 2011). However, ‘camouflaging’ is said to be a central characteristic of the female phenotype of ASC which account for the gender bias in diagnosis (Hull et

al., 2017; Lai et al., 2015). This does not then mean that females with ASC do not need support with social interactions or facial emotion recognition. It may be that Mind Reading is effective for use in female children or it may need to be adapted. Future research could focus on implementing Mind Reading in the UK and examining different participant characteristics such as socioeconomic status. Moreover, the research could specifically aim to recruit female participants with ASC.

Conclusion and Recommendations

Previously noted limitations of interventions that develop emotion recognition in children with ASC include that they require considerable time, resources, coordination, and expertise to implement (Lord et al., 2005; Rao et al., 2008). However, the computerized format of Mind Reading mitigates these limitations, for example the programme administers itself and does not require any additional facilitator. In addition, the cost of the programme is low making Mind Reading much more feasible than other forms of emotion recognition intervention. For this reason, Mind Reading has potential applications for children with ASC in schools (such as mainstream or special schools) and at home. Young people spend a substantial amount of their time in school surrounded by others. Ostensibly children with difficulty in emotional recognition find social interaction challenging. The school environment provides an appropriate opportunity to develop children with ASC's emotion recognition skills and Mind Reading appears to be a suitable intervention to implement. Delineating Mind Reading's effectiveness in real-life settings will be an important area for future work.

Reference List

American Psychiatric Association (2013). Diagnostic and statistical manual of mental disorders. 5th ed. Arlington, VA: American Psychiatric Association

American Psychiatric Association. (1994). Diagnostic and Statistical Manual of Mental Disorders (4th ed.).

Ashwin, C., Chapman, E., Colle, L., & Baron-Cohen, S. (2006). Impaired recognition of negative basic emotions in autism: A test of the amygdala theory. *Social Neuroscience*, *1*(3-4), 349-363. doi: 10.1080/17470910601040772

Bal, E., Harden, E., Lamb, D., Van Hecke, A. V., Denver, J. W., & Porges, S. W. (2010). Emotion recognition in children with autism spectrum disorders: Relations to eye gaze and autonomic state. *Journal of autism and developmental disorders*, *40*(3), 358-370.

Barker, Chris & Pistrang, Nancy & Elliott, Robert. (2016). Research methods in clinical psychology: An introduction for students and practitioners (3rd ed.). 10.1002/9781119154082.

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends In Cognitive Sciences*, *6*(6), 248-254. doi: 10.1016/s1364-6613(02)01904-6

Baron-Cohen, S. (2003). *The essential difference: The truth about the male and female brain.* Basic Books.

Baron-Cohen, S., Golan, O., & Ashwin, E. (2009). Can emotion recognition be taught to children with autism spectrum conditions?. *Philosophical Transactions Of The Royal Society B: Biological Sciences*, *364*(1535), 3567-3574. doi: 10.1098/rstb.2009.0191

Bishop, J. (2003). The Internet for educating individuals with social impairments. *Journal of Computer Assisted Learning*, *19*(4), 546-556.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, N.J.: L. Erlbaum Associates.

Constantino, J. N., & Gruber, C. P. (2012). *Social responsiveness scale: SRS-2*. Torrance, CA: Western psychological services.

Davidson, D., Hilvert, E., Winning, A. M., & Giordano, M. (2021). Recognition of Emotions from Situational Contexts and the Impact of a Mind Reading Intervention in Children with Autism Spectrum Disorder. *Child Psychiatry and Human Development*. https://doi.org/10.1007/s10578-021-01139-0

Department for Education (2017). Transforming Children and Young People's Mental Health Provision: A Green Paper. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664855/Transforming_children_and_young_people_s_mental_health_provision.pdf

Department for Education (2019). Retrieved 21 June 2022, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/814244/SEN_2019_Text.docx.pdf

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. Psychological Methods, 1(2), 170–177. https://doi.org/10.1037/1082-989X.1.2.170

Ebbinghaus, H. (1964). *Memory*. Рипол Классик.

Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental psychology and nonverbal behavior*, *1*(1), 56-75.

Fombonne, E. (2009). Epidemiology of pervasive developmental disorders. *Pediatric research*, *65*(6), 591-598.

Fonagy, P., Cottrell, D., Phillips, J., Bevington, D., Glaser, D., & Allison, E. (2014). What works for whom?: a critical review of treatments for children and adolescents.

Gersten, R., Fuchs, L., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. (2005). Quality Indicators for Group Experimental and Quasi-Experimental Research in Special Education. *Exceptional Children*, *71*(2), 149-164. doi: 10.1177/001440290507100202

Golan, O., Sinai-Gavrilov, Y., & Baron-Cohen, S. (2015). The Cambridge Mindreading Face-Voice Battery for Children (CAM-C): complex emotion recognition in children with and without autism spectrum conditions. *Molecular Autism*, *6*(1). doi: 10.1186/s13229-015-0018-z

Gough, D. (2007). Weight of Evidence: a framework for the appraisal of the quality and relevance of evidence. Research Papers In Education, 22(2), 213-228. doi: 10.1080/02671520701296189

Gross, T. F. (2004). The perception of four basic emotions in human and nonhuman faces by children with autism and other developmental disabilities. *Journal of abnormal child psychology*, *32*(5), 469-480

Harden, A., & Gough, D. (2012). Quality and relevance appraisal. Sage Publications.

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.

Heck, A., Chroust, A., White, H., Jubran, R., & Bhatt, R. S. (2018). Development of body emotion perception in infancy: From discrimination to recognition. *Infant Behavior and Development*, *50*, 42-51.

Higgins, J., & Thompson, S. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics In Medicine*, *21*(11), 1539-1558. doi: 10.1002/sim.1186

Horner, R., Carr, E., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The Use of Single-Subject Research to Identify Evidence-Based Practice in Special Education. *Exceptional Children*, *71*(2), 165-179. doi: 10.1177/001440290507100203

Hull, L., Petrides, K. V., Allison, C., Smith, P., Baron-Cohen, S., Lai, M. C., & Mandy, W. (2017). "Putting on my best normal": Social camouflaging in adults with autism spectrum conditions. *Journal of autism and developmental disorders*, *47*(8), 2519-2534

Iordanou, C., & Mattock, K. (2021). Where the Wild Things Are: understanding of emotions in a picture book. *Education 3-13*, 1-13. doi: 10.1080/03004279.2021.1882526

Karpicke, J. D., & Roediger III, H. L. (2008). The critical importance of retrieval for learning. *science*, *319*(5865), 966-968.

Karpicke, J. D., Blunt, J. R., & Smith, M. A. (2016). Retrieval-based learning: Positive effects of retrieval practice in elementary school children. *Frontiers in Psychology*, *7*, 350.

Lacava, P. G., Rankin, A., Mahlios, E., Cook, K., & Simpson, R. L. (2010). A single case design evaluation of a software and tutor intervention addressing emotion recognition and social interaction in four boys with ASD. *Autism*, *14*(3), 161–178. https://doi.org/10.1177/1362361310362085

Lacava, P., Golan, O., Baron-Cohen, S., & Smith Myles, B. (2007). *Using Assistive Technology to Teach Emotion Recognition to Students With Asperger Syndrome*. 174–181. https://doi.org/10.1177/07419325070280030601

Lai, M. C., Lombardo, M. V., Auyeung, B., Chakrabarti, B., & Baron-Cohen, S. (2015). Sex/gender differences and autism: setting the scene for future research. *Journal of the American Academy of Child & Adolescent Psychiatry*, *54*(1), 11-24.

Lipsey, M., & Wilson, D. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. American Psychologist, 48(12), 1181-1209. doi: 10.1037/0003-066x.48.12.1181

Lopata, C., Thomeer, M. L., Rodgers, J. D., Donnelly, J. P., & McDonald, C. A. (2016). RCT of mind reading as a component of a psychosocial treatment for high-functioning children with ASD. *Research in Autism Spectrum Disorders*, *21*, 25–36. https://doi.org/10.1016/j.rasd.2015.09.003

Lopata, C., Thomeer, M. L., Rodgers, J. D., Donnelly, J. P., McDonald, C. A., Volker, M. A., Smith, T. H., & Wang, H. (2019). Cluster Randomized Trial of a School Intervention for Children with Autism Spectrum Disorder. *Journal of Clinical Child and Adolescent Psychology*, *48*(6), 922–933. https://doi.org/10.1080/15374416.2018.1520121

Lopata, C., Thomeer, M. L., Volker, M. A., Lee, G. K., Smith, T. H., Rodgers, J. D., Smith, R. A., Gullo, G., McDonald, C. A., Mirwis, J., & Toomey, J. A. (2013). Open-Trial Pilot Study of a Comprehensive School-Based Intervention for High-Functioning Autism Spectrum Disorders. *Remedial and Special Education*, *34*(5), 269–281. https://doi.org/10.1177/0741932512450518

Lopata, C., Thomeer, M. L., Volker, M. A., Lee, G. K., Smith, T. H., Smith, R. A., Mcdonald, C. A., Rodgers, J. D., Lipinski, A. M., & Toomey, J. A. (2012). Feasibility and Initial Efficacy of a Comprehensive School-Based Intervention for

High-Functioning Autism Spectrum Disorders. *Psychology in the Schools*, *49*(10), 963–974. https://doi.org/10.1002/pits.21649

Lopata, C., Thomeer, M. L., Volker, M. A., Nida, R. E., & Lee, G. K. (2008). Effectiveness of a manualized summer social treatment program for high-functioning children with autism spectrum disorders. *Journal of autism and developmental disorders*, *38*(5), 890-904.

Lord, C., Wagner, A., Rogers, S., Szatmari, P., Aman, M., & Charman, T. et al. (2005). Challenges in Evaluating Psychosocial Interventions for Autistic Spectrum Disorders. Journal Of Autism And Developmental Disorders, 35(6), 695-708. doi: 10.1007/s10803-005-0017-6

Marsh, A. A., & Blair, R. J. R. (2008). Deficits in facial affect recognition among antisocial populations: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, *32*(3), 454-465.

Moore, D., McGrath, P., & Thorpe, J. (2000). Computer-aided learning for people with autism–a framework for research and development. *Innovations in education and training international*, *37*(3), 218-228.

Nind, M., and Wearmouth, J. (2005) A systematic review of pedagogical approaches that can effectively include children with special educational needs in mainstream classrooms with a particular focus on peer group interactive approaches London, UK. EPPI-Centre Research Evidence in Education Library

Paynter, J., & Peterson, C. (2013). Further evidence of benefits of thought-bubble training for theory of mind development in children with autism spectrum disorders. Research In Autism Spectrum Disorders, 7(2), 344-348. doi: 10.1016/j.rasd.2012.10.001

Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: horses for courses. Journal Of Epidemiology &Amp; Community Health, 57(7), 527-529. doi: 10.1136/jech.57.7.527

Petticrew, M., & Roberts, H. (2006). Systematic reviews in the social sciences: A practical guide. Blackwell Publishing. https://doi.org/10.1002/9780470754887

Rao, P., Beidel, D., & Murray, M. (2007). Social Skills Interventions for Children with Asperger's Syndrome or High-Functioning Autism: A Review and Recommendations. Journal Of Autism And Developmental Disorders, 38(2), 353-361. doi: 10.1007/s10803-007-0402-4

Reynolds, C. R., Kamphaus, R. W. (2004). Behavior Assessment System for Children–Second Edition (BASC-2). Bloomington, MN: Pearson.

Russell, G., Steer, C., & Golding, J. (2011). Social and demographic factors that influence the diagnosis of autistic spectrum disorders. *Social psychiatry and psychiatric epidemiology*, *46*(12), 1283-1293.

Schultz, R. T. (2005). Developmental deficits in social perception in autism: the role of the amygdala and fusiform face area. *International Journal of Developmental Neuroscience*, *23*(2-3), 125-141

Sodian, B., & Thoermer, C. (2008). Precursors to a theory of mind in infancy: Perspectives for research on autism. *Quarterly Journal of Experimental Psychology*, *61*(1), 27-39.

Suurmond, R., van Rhee, H., & Hak, T. (2017). Introduction, comparison, and validation ofMeta-Essentials: A free and simple tool for meta-analysis. *Research Synthesis Methods*, *8*(4), 537-553. doi: 10.1002/jrsm.1260

Tardif, C., Lainé, F., Rodriguez, M., & Gepner, B. (2007). Slowing down presentation of facial movements and vocal sounds enhances facial expression recognition

and induces facial–vocal imitation in children with autism. *Journal of autism and developmental disorders*, *37*(8), 1469-1484

Thomeer, M. L., Smith, R. A., Lopata, C., Volker, M. A., Lipinski, A. M., Rodgers, J. D., McDonald, C. A., & Lee, G. K. (2015). Randomized Controlled Trial of Mind Reading and In Vivo Rehearsal for High-Functioning Children with ASD. *Journal of Autism and Developmental Disorders*, *45*(7), 2115–2127. https://doi.org/10.1007/s10803-015-2374-0

Thomeer, M., Rodgers, J., Lopata, C., McDonald, C., Volker, M., & Toomey, J. et al. (2011). Open-Trial Pilot of Mind Reading and In Vivo Rehearsal for Children With HFASD. Focus On Autism And Other Developmental Disabilities, 26(3), 153-161. doi: 10.1177/1088357611414876

Weinger, P. M., & Depue, R. A. (2011). Remediation of deficits in recognition of facial emotions in children with Autism spectrum disorders. *Child and Family Behavior Therapy*, *33*(1), 20–31. https://doi.org/10.1080/07317107.2011.545008

Whiffen, J. W., & Karpicke, J. D. (2017). The role of episodic context in retrieval practice effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(7), 1036.

Widen, S., & Russell, J. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Developmental Psychology*, *39*(1), 114-128. doi: 10.1037/0012-1649.39.1.114

Wigelsworth, M., Verity, L., Mason, C-J., Humphrey, N., Qualter, P., & Troncoso, P. (2020). Programmes to practices: Identifying effective, evidence-based social and emotional learning strategies for teachers and schools. Education Endowment Foundation.

https://educationendowmentfoundation.org.uk/public/files/Social_and_Emotional_Learning_Evidence_Review.pdf

Golan, Sinai-Gavrilov White, S., Keonig, K., & Scahill, L. (2007). Social Skills Development in Children with Autism Spectrum Disorders: A Review of the Intervention Research. Journal Of Autism And Developmental Disorders, 37(10), 1858-1868. doi: 10.1007/s10803-006-0320-x

Appendices

Appendix A: Mapping Table

Author	Country	Participants	Study Design	Intervention	Measures	Outcomes
Davidson et al. (2021 Study 2)	United States	24 aged 6-9	Quasi-experimental	Engaged in study that used Situational Emotion Task Measure then engaged in Mind Reading	Pre and post number of correctly identified emotions	The repeated measures ANOVA revealed main effects of Time of Testing, $F(1, 17) = 4.86$, $p = 0.042$, $\eta_p^2 = 0.22$, Emotion, $F(5, 85) = 3.55$, $p = 0.006$, $\eta_p^2 = 0.17$, and Intensity of Emotion, $F(1, 17) = 14.41$, $p = 0.001$, $\eta_p^2 = 0.46$. These findings were qualified by an Emotion X Time of Testing two-way interaction, $F(5, 85) = 2.35$, $p = 0.048$, $\eta_p^2 = 0.12$, and an Emotion X Intensity of Emotion interaction, $F(5, 85) = 4.47$, $p = 0.001$, $\eta_p^2 = 0.21$. For the Emotion X Time of Testing interaction, post-hoc analyses revealed significant improvements for recognizing guilt, $t(17) = -3.92$, $p = 0.001$, and embarrassment, $t(17) = -2.92$, $p = 0.004$, from pre- to post-intervention. For the Emotion X Intensity of Emotion interaction, post-hoc analyses revealed that high intensity examples of guilt, $t(17) = 3.75$, $p = 0.012$, and embarrassment, $t(17) = -2.76$, $p = 0.014$, were more accurately recognized by children than low intensity examples following the intervention.
Davidson, D., Hilvert, E., Winning, A. M., & Giordano, M. (2021). Recognition of Emotions from Situational Contexts and the Impact of a Mind Reading Intervention in Children with Autism Spectrum Disorder. *Child Psychiatry and Human Development*. https://doi.org/10.1007/s10578-021-01139-0						
Author	Country	Participants	Study Design	Intervention	Measures	Outcomes
Lopata et al. (2013)	United States	24 aged 6-9	Quasi-experimental single-group open trial	Mind Reading as part of overall comprehensive school-based interventions (CSBIs) for children with high-functioning autism spectrum	The CAM-C DANVA2 BASC-2	Large effect sizes ($d \geq .80$) were obtained on recognition of emotional expressions taught in treatment (CAM-C), and parent ratings of broader social skills (BASC-2). Medium effect sizes ($d \geq .50$) were obtained for a child test of broader emotion recognition skills (DANVA2).
Lopata, C., Thomeer, M. L., Volker, M. A., Lee, G. K., Smith, T. H., Rodgers, J. D., Smith, R. A., Gullo, G., McDonald, C. A., Mirwis, J., & Toomey, J. A. (2013). Open-Trial Pilot Study of a Comprehensive School-Based Intervention for High-Functioning Autism Spectrum Disorders. *Remedial and Special Education*, *34*(5), 269–281. https://doi.org/10.1177/0741932512450518						
Author	Country	Participants	Study Design	Intervention	Measures	Outcomes
Lopata et al., 2019)	United States	103 aged 6–12 52 in intervention group, 51 in control	Cluster Randomized Trial	Mind Reading part of wider comprehensive school-based intervention called schoolMAX	CAM-C SRS-2	Results for the CAM-C revealed a significant treatment effect, $F(1, 100) = 33.16$, $p < .001$, $d = 1.41$, CI [.74, 2.09] (school ICC = .28); children in the intervention group demonstrated a significantly greater increase compared to those in the control condition. Similarly, parent–teacher SRS-2 ratings showed that children in the intervention had a significantly greater decrease in ASC symptoms compared to children in the control group, $F(1, 100) = 23.51$, $p < .001$, $d = -1.15$, CI [−1.77, −.53] (school ICC = .22).
Lopata, C., Thomeer, M. L., Rodgers, J. D., Donnelly, J. P., McDonald, C. A., Volker, M. A., Smith, T. H., & Wang, H. (2019). Cluster Randomized Trial of a School Intervention for Children with Autism Spectrum Disorder. *Journal of Clinical Child and Adolescent Psychology*, *48*(6), 922–933. https://doi.org/10.1080/15374416.2018.1520121						
Author	Country	Participants	Study Design	Intervention	Measures	Outcomes
Lopata et al. (2016)	United States	36 aged 7–12 18 in intervention group 18 in control group	Randomized control trial	Mind Reading as part of wider comprehensive school-based intervention called summerMAX	CAM-C ERDS SEE ASC BASC-2 SRS	On the CAM-C Faces child test, a significant time × treatment condition interaction was observed ($p = .003$; large effect) favoring the group that engaged in Mind Reading. Significant main effects for time were found parent and clinician ratings on the ERDS (ps < .001). The effect sizes on each of these main effects were large.
Lopata, C., Thomeer, M. L., Rodgers, J. D., Donnelly, J. P., & McDonald, C. A. (2016). RCT of Mind Reading as a component of a psychosocial treatment for high-functioning children with ASD. *Research in Autism Spectrum Disorders*, *21*, 25–36. https://doi.org/10.1016/j.rasd.2015.09.003						

Author	Country	Participants	Study Design	Intervention	Measures	Outcomes
Lopata et al. (2012)	United States	12 aged 6-9	Quasi-experimental	Mind Reading	SKA CAM-C DANVA2 *SSC.* *BASC-2.*	Large effect sizes ($d \geq .80$) were obtained on SKA, (CAM-C, and broader emotion recognition skills in children's faces DANVA2 and parent ratings of skills taught in the curriculum ASC and autism symptoms. Medium effect sizes ($d \geq .50$) were obtained for teacher ratings of skills taught in the program ASC and teacher and parent ratings of broader social skills BASC-2.

Lopata, C., Thomeer, M. L., Volker, M. A., Lee, G. K., Smith, T. H., Smith, R. A., Mcdonald, C. A., Rodgers, J. D., Lipinski, A. M., & Toomey, J. A. (2012). Feasibility and Initial Efficacy of a Comprehensive School-Based Intervention for High-Functioning Autism Spectrum Disorders. *Psychology in the Schools*, *49*(10), 963–974. https://doi.org/10.1002/pits.21649

Author	Country	Participants	Study Design	Intervention	Measures	Outcomes
Thomeer et al. (2011)	United states	11 aged 7-12	Quasi-experimental open trial pilot	In vivo rehearsal and Mind Reading	SRS ERDS	Pre–post comparisons were conducted for EDRS. Significant pre versus post differences were found in decoding and encoding. Examination of mean scores indicated significantly higher posttest scores (compared with pretest) for decoding and encoding of emotions, with effect sizes falling in the large range.

Thomeer, M., Rodgers, J., Lopata, C., McDonald, C., Volker, M., & Toomey, J. et al. (2011). Open-Trial Pilot of Mind Reading and In Vivo Rehearsal for Children With HFASD. Focus On Autism And Other Developmental Disabilities, 26(3), 153-161. doi: 10.1177/1088357611414876

Author	Country	Participants	Study Design	Intervention	Measures	Outcomes
Thomeer et al. (2015)	United States	43 aged 7-12 22 assigned to intervention group 21 assigned to control group	Randomized Controlled Trial	Mind Reading and in vivo rehearsal	CAM-C ERDS SRS BASC-2	Results of the ANCOVA for CAM-C Faces yielded a significant between-groups effect (p <.001; $[\text{omega}]^2$ =.23). *Post hoc* comparisons between the two conditions indicated that the intervention group achieved a significantly higher CAM-C Faces score than the control group at both posttest (t [40] = 5.79, p <.001 [one-tail], d = 1.34) and follow-up (t [40] = 3.45, p =.001 [one-tail], d = 0.86). Between-groups effect size estimates at posttest and follow-up were large. For the ERDS Expressive (encoding) ratings, ANCOVA results were significant (p =.0025, $[\text{omega}]^2$ =.11). *Post hoc* between-groups differences were significant and favoured the intervention group at posttest (t [40] = 2.33, p =.0125 [one-tail], d =.61) and at follow-up (t [40] = 2.93, p =.003 [one-tail], d =.85). Between-groups effect size estimates were medium at posttest and large at follow-up.

Thomeer, M. L., Smith, R. A., Lopata, C., Volker, M. A., Lipinski, A. M., Rodgers, J. D., McDonald, C. A., & Lee, G. K. (2015). Randomized Controlled Trial of Mind Reading and In Vivo Rehearsal for High-Functioning Children with ASD. *Journal of Autism and Developmental Disorders*, *45*(7), 2115–2127. https://doi.org/10.1007/s10803-015-2374-0

Author	Country	Participants	Study Design	Intervention	Measures	Outcomes
Weinger & Depue, (2011)	United States	6 aged 7-11	Quasi-experimental	Mind Reading	Pre and post number of correctly identified emotions	The Mind Reading computer software appears to significantly ($p < .001$) improve the emotion recognition abilities in children with ASD. With training, participants in the clinical sample reached levels of ER comparable to the levels of ER observed in our age-matched control sample (both groups having a mean of approximately 90%).

Weinger, P. M., & Depue, R. A. (2011). Remediation of deficits in recognition of facial emotions in children with Autism spectrum disorders. *Child and Family Behavior Therapy*, *33*(1), 20–31. https://doi.org/10.1080/07317107.2011.545008

Author	Country	Participants	Study Design	Intervention	Measures	Outcomes
Lacava et al. (2007)	United States	8 aged 8-11	Quasi-experimental	Mind Reading	CAM-C C-FAT RMF-C	The differences between the pre- and posttest performances on all three measures were statistically significant for all tasks, with mean posttest scores higher than pretest scores. This was found for the CAM-C Faces subtest, $z = -2.366$, $p < .05$; and the C-FAT, $z = -2.028$, $p < .05$. Participants' performance on the RMF-C task resulted in a mean score of 13.375 (SD = 4.172). While this average score appears lower than that of Golan's (2006) groups of children with ASC who received no intervention (M = 14.52, SD = 3.61), and the

						group of children with ASC who used Mind Reading (M = 15.48, SD = 2.54), these differences were not statistically significant.
Lacava, P., Golan, O., Baron-Cohen, S., & Smith Myles, B. (2007). *Using Assistive Technology to Teach Emotion Recognition to Students With Asperger Syndrome*. 174–181. https://doi.org/10.1177/07419325070280030601						
Author	Country	Participants	Study Design	Intervention	Measures	Outcomes
Lacava et al. (2010)	United States	4	Single case design	Mind Reading and tutor intervention	CAM-C	All participants improved emotion recognition tests scores from pre- to post-testing. On the CAM-C, all participants increased their scores. All participants made improvements in both basic and complex emotion recognition.
Lacava, P. G., Rankin, A., Mahlios, E., Cook, K., & Simpson, R. L. (2010). A single case design evaluation of a software and tutor intervention addressing emotion recognition and social interaction in four boys with ASD. *Autism*, *14*(3), 161–178. https://doi.org/10.1177/1362361310362085						

Appendix B: List of excluded studies following abstract screening

Article	Exclusion criteria number(s)
Lopata, C., Donnelly, J., Thomeer, M., Rodgers, J., Lodi-Smith, J., Booth, A., & Volker, M. (2020). Moderators of School Intervention Outcomes for Children with Autism Spectrum Disorder. *Journal Of Abnormal Child Psychology*, *48*(8), 1105-1114. doi: 10.1007/s10802-020-00652-5	5 This study did not employ Mind Reading as an intervention. In this study, a range of demographic, clinical, and school variables were tested as potential moderators of treatment outcomes. The data was drawn from a study that had been conducted previously.
Williams, B., Gray, K., & Tonge, B. (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. doi: 10.1111/j.1469-7610.2012.02593.x	5 This study did not employ Mind Reading as an intervention. The study taught emotion recognition through the use of a DVD 'The Transporters' (Changing Media Development, 2006)[1]
Bernard-Opitz, V., Sriram, N., & Nakhoda-Sapuan, S. (2001). *Journal Of Autism And Developmental Disorders*, *31*(4), 377-384. doi: 10.1023/a:1010660502130	5 This study did not employ Mind Reading as an intervention. Participants asked to provide solutions to animated problem scenes
Bölte, S., Feineis-Matthews, S., Leber, S., Dierks, T., Hubl, D., & Poustka, F. (2002). The development and evaluation of a computer-based program to test and to teach the recognition of facial affect. *International Journal Of Circumpolar Health*, *61*(0). doi: 10.3402/ijch.v61i0.17503	5 This study did not employ Mind Reading as an intervention. The authors developed and investigated their own intervention.
Faja, S., Aylward, E., Bernier, R., & Dawson, G. (2007). Becoming a Face Expert: A Computerized Face-Training Program for High-Functioning Individuals With Autism Spectrum Disorders. *Developmental Neuropsychology*, *33*(1), 1-24. doi: 10.1080/87565640701729573	5 This study did not employ Mind Reading as an intervention. The study investigated the use of an intervention based on a modified version of Gauthier and Tarr's (1997) [2] expertise protocol

[1] Changing Media Development (2006). The Transporters. London, UK: Changing Media Development Limited.

[2] Gauthier, I., & Tarr, M. J. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. Vision Research, 37(12), 1673–1682. https://doi.org/10.1016/S0042-6989(96)00286-6

Russo-Ponsaran, N., Evans-Smith, B., Johnson, J., Russo, J., & McKown, C. (2015). Efficacy of a Facial Emotion Training Program for Children and Adolescents with Autism Spectrum Disorders. *Journal Of Nonverbal Behavior*, *40*(1), 13-38. doi: 10.1007/s10919-015-0217-5	5 This study did not employed Mind Reading as an intervention. The intervention of interest in this study was a modification of a commercially-available, computerized, dynamic facial emotion training tool, the MiX by Humintell©[3]

[3] Matsumoto, D., & Hwang, H. S. (2011). Evidence for training the ability to read microexpressions of emotion. Motivation and Emotion, 35, 181–191. doi:10.1007/s11031-011-9212-2

Appendix C: WoE A

Gersten et. al. (2005) 'Quality indicators for group experimental and quasi-experimental research in special education' coding protocol presents quality indicators for experimental and quasi-experimental studies for special education. The indicators are intended to be used as an organizer of critical issues for consideration in research and to evaluate the merits of a completed research report or article. Gersten et. al. (2005, p.1) 'believe these indicators can be used widely, from assisting in the development of research plans to evaluating proposals'. The coding protocol includes both essential and desirable criteria. Gersten et al. (2005) coding protocol is employed to evaluate the description of participants, procedure for implementing the intervention, outcome measures used and how the data collected was analysed.

See sample below:

Coding protocol: Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C, & Innocenti, M. (2005). Quality indicators for group experimental and quasi-experimental research in special education. Exceptional Children, 71,149-164.	
Authors: Davidson, et al.(2021) Study 2	
Essential Quality Indicators	
Quality Indicators for Describing Participants	
1 .Was sufficient information provided to determine/confirm whether the participants demonstrated any disability(ies) or difficulties?	Yes
2 .Were appropriate procedures used to increase the likelihood that relevant characteristics of participants in the sample were comparable across conditions?	NA
3 .Was sufficient information given characterizing the role of the teachers or researchers provided?	Yes

4. Was sufficient information given to indicate if the role of the teachers or researchers was comparable across conditions?	NA
Quality Indicators for Implementation of the Intervention and Description of Comparison Conditions	
5 . Was the intervention clearly described and specified?	Yes
6 . Was the fidelity of implementation described and assessed?	Yes
7 . Was the nature of services provided in comparison conditions described?	NA
Quality Indicators for Outcome Measures	
8 8. Were multiple measures used to provide an appropriate balance between measures closely aligned with the intervention and measures of generalized performance?	No
9. Were outcomes for capturing the interventions effect measured at the appropriate times?	Yes
Quality Indicators for Data Analysis	
10 Were the data analysis techniques appropriately linked to key research questions and hypotheses?	Yes
11 Were the data analysis techniques appropriately linked to the unit of analysis in the study?	Yes
12 Did the research report include not only inferential statistics but also effect size calculations?	Yes
Number of indicators met out of 12	8
Desirable Quality Indicators	
1 Where missing test results reported in the final outcomes?	No

2	Did the study provide internal consistency reliability but also test- retest reliability?	No
3	Did the study report interrater reliability (when appropriate) for outcome measures?	NA
4	Were data collectors and/or scorers blind to study conditions and equally (un)familiar to examinees across study conditions?	NA
5	Were outcomes for capturing the intervention's effect measured beyond an immediate post-test?	No
6	Was evidence of the criterion-related validity and construct validity of the measures provided?	No
7	Did the research team assess surface features of fidelity implementation (e.g., number of minutes allocated to the intervention or teacher/interventionist following procedures specified)?	Yes
8	Did the research team assess quality of implementation?	No
9	Was any documentation of the nature of instruction or series provided in comparison conditions?	NA
10	Did the research report include illustrations that capture the nature of the intervention?	Yes
11	Were results presented in a clear, coherent fashion?	Yes
Number of indicators met out of 11		3
Essential Quality Indicators Less than 9 = Score 0 Greater than 9 = Score 1		0
Desirable Quality Indicators Total of 0 = 0		1

Less than 4 = Score 1 Greater than 4 = Score 2	
Total Quality Score	1

Coding Protocol: Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M.(2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, 71(2), 165-179)
Lacava, P. G., Rankin, A., Mahlios, E., Cook, K., & Simpson, R. L. (2010). A single case design evaluation of a software and tutor intervention addressing emotion recognition and social interaction in four boys with ASD

Section A: Description of Participants and Setting	
1. Participants are described with sufficient detail to allow others to select individuals with similar characteristics; (e.g., age, gender, disability, diagnosis).	Yes
2. The process for selecting participants is described with replicable precision.	Yes
3. Critical features of the physical setting are described with sufficient precision to allow replication	Yes
Section A score:	3
Section B: Dependent Variable	
1. Dependent variables are described with operational precision	Yes
2. Each dependent variable is measured with a procedure that generates a quantifiable index	No
3. Measurement of the dependent variable is valid and described with replicable precision	No
Dependent variables are measured repeatedly over time	No
Data are collected on the reliability or inter-observer agreement associated with each dependent variable, and IOA levels meet minimal standards {e.g., IOA = 80%; Kappa = 60%).	Yes
Section B score:	2
Section C: Independent Variable	
Independent variable is described with replicable precision	Yes
Independent variable is systematically manipulated and under the control of the experimenter	No
Overt measurement of the fidelity of implementation for the independent variable is highly desirable.	Yes
Section C score:	2
Section D: Baseline	

The majority of single-subject research studies will include a baseline phase that provides repealed measurement of a dependent variable and establishes a pattern of responding that can be used to predict the pattern of future performance if introduction or manipulation of the independent variable did not occur	No
Baseline conditions are described with replicable precision	No
Section D score:	0
Section E: Experimental Control/ Internal Validity	
The design provides at least three demonstrations of experimental effect at three different pointsin time.	Yes
The design controls for common threats to internal validity (e.g., permits elimination of rival hypotheses).	No
The results document a pattern that demonstrates experimental control	No
Section E score:	1
Section F: External Validity	
Experimental effects are replicated across participants, settings, or materials to establish external validity.	Yes
Section F score:	1
Section G: Social Validity	
The dependent variable is socially important.	Yes
The magnitude of change in the dependent variable resulting from the intervention is socially important.	Yes
Implementation of the independent variable is practical and cost effective.	Yes
Social validity is enhanced by implementation of the independent variable over extended time periods, by typical intervention agents, in typical physical and social contexts.	No
Section G score:	3
Guide for weighting Low equal to or less than 1.4, Medium 1.5 to 2.4 and High more than 2.5	
Overall weight of evidence rating	1.7 Medium

Weight of Evidence (WoE) Calculations

Section	Overall Evidence Rating (0 – 3)	Evidence Descriptors

A: Description of Participants and Settings	3	High
B: Dependent Variable	2	Medium
C: Independent Variable	2	Medium
D: Baseline	0	Low
E: Experimental control/internal validity	1	Low
F: External Validity	1	Low
G: Social Validity	3	High
Note. <1.5 is low; 1.5 – 2.4 is medium; >2.4 is high		

How to calculate Average Quality of Evidence across the Key Judgement Areas

Equation: $\frac{\sum x}{N}$

x = Individual quality score for each section

N = Numer of judgement areas (7)

Overall weight of evidence rating = 1.7 (Medium)

Appendix D

Weight of Evidence B: Appropriateness of Design
The weight of evidence B was evaluated using the hierarchy from Petticrew and Roberts (2003) who stated that in order to answer a question of 'effectiveness' the studies most appropriate to use should follow this system of weighting. Table C 3 displays WoE B criteria: (systematic literature reviews have been removed from the list). Table C 4 displays the judgments made for the studies included in this review.

See sample below

Weight of Evidence B: Appropriateness of Design	
Weighting	Description
High (3)	1. Randomised control trials
Medium (2)	2. Cohort studies, quasi-experimental studies, single case experimental designs
Low (1)	3. Qualitative research, survey, case-control, nonexperimental evaluation

Authors	Weight of Evidence B: Appropriateness of Design weighting
Davidson et al.(2021,Study 2)	Medium (2)

Appendix E

WoE C is a review-specific judgement about how relevant the focus of study is to the review question. The tables below illustrates the criteria by which each study was assessed and the WoE C rating

WoE C Criteria and Rationale

Criteria	WoE Rating	Descriptor	Rationale
Includes CAM-C [4]and broader measures such as those completed by parents/teachers (for example SRS-2 [5]or BASC-2[6])	1	Did not use CAM-C	The CAM-C is an adaptation of a complex emotion recognition battery for adults and is specifically employed within the Mind Reading programme to assess facial emotion recognition of children with ASC. Presently it has been reported that there is a lack of measures that evaluate outcomes of interventions, sensitive to the core features of ASCs. Recommendations to improve assessment of outcomes in ASCs include the use of multiple instruments that measure both specific and general areas of performance (Lord et al., 2005; Williams et al., 2007) and to allow ratings by multiple informants (Williams et al 2007). Which is why this criterion includes a high rating for the use of CAM-C and additional measures completed by a range of informants.
	2	Used CAM-C	
	3	Used additional measures of only the child's performance	
	4	Used additional measure completed by parents/teachers	
Examines Mind Reading exclusively	1	Used concurrent interventions	This review is looking at the effectiveness of a specific intervention. Including Mind Reading exclusively, reduces the possibility that conclusions drawn about the intervention are influences by the use of any additional interventions.
	2	Used other intervention preceding the implementation of Mind Reading	
	3	Used Mind Reading only	
Setting	1	Delivered in research setting university/college campus	To improve external validity, it is preferred for the study to be conducted in a naturalistic setting. As educational psychologists often work within education settings, setting such as schools received a higher rating.
	2	Delivered at home	

3	Delivered in school	

[4] [4] The Cambridge Mindreading Face-Voice Battery for Children (CAM-C; Golan, et al., 2015)

[5] Social responsiveness scale 2 (SRS-2; Constantino & Gruber, 2012)

[6] The Cambridge Mindreading Face-Voice Battery for Children (CAM-C; Golan, et al., 2015)

Weight of Evidence C

See sample below

		Author: Davidson, et al.(2021) Study 2
	Score	Rating
Includes CAM-C and broader measures such as those completed by parents/teachers (for example SRS-2 or BASC-2)	1 Did not use CAM-C	1
	2 Used CAM-C	
	3 Used additional measures of only the child's performance	
	4 Used additional measure completed by parents/teachers	
Examines Mindreading exclusively	1 Used concurrent interventions	2
	2 Used other intervention preceding the implementation of Mind Reading	
	3 Used Mind Reading only	
Setting	1 Delivered in research setting university/college campus	3
	2 Delivered at home	
	3 Delivered in school	
Total		6
Weighting		Med

Guide for weighting Low – 1- 3, Medium 4 - 6 and High 7 - 10

Appendix F

Weight of Evidence D (WoE D): Overall WoE

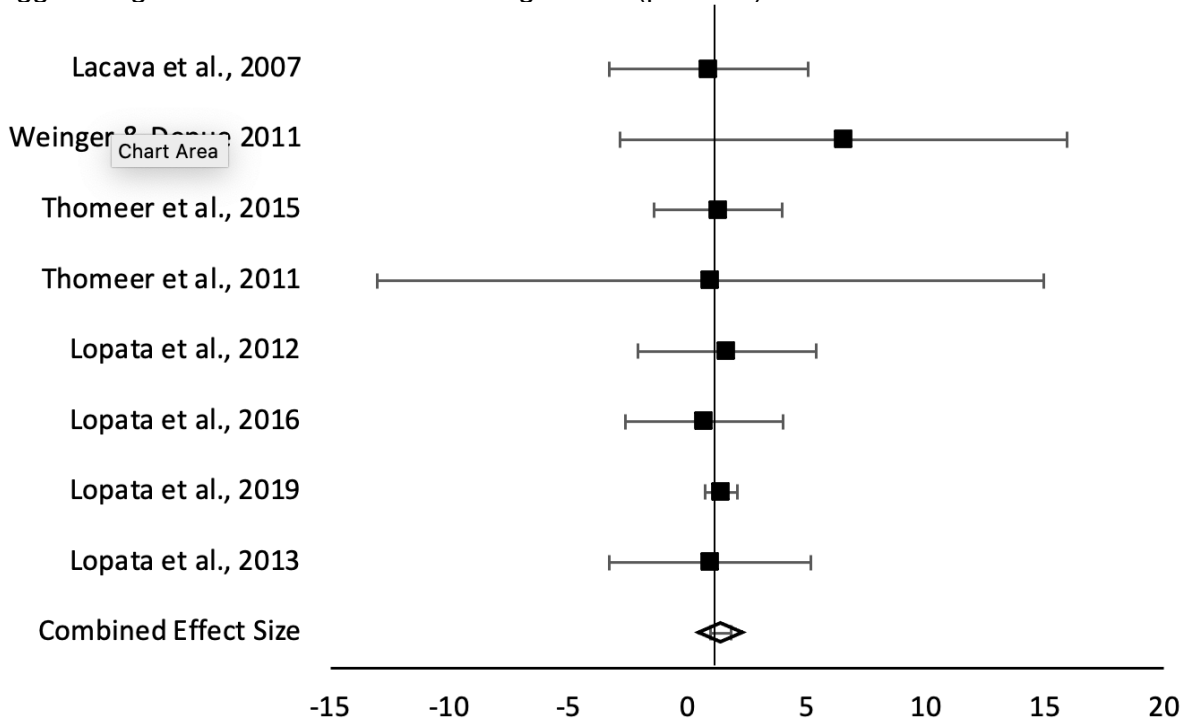
See sample below

Score	WoE D Descriptor
Less than 2	Low
2 to 2.5	Medium
Greater than 2.5	High

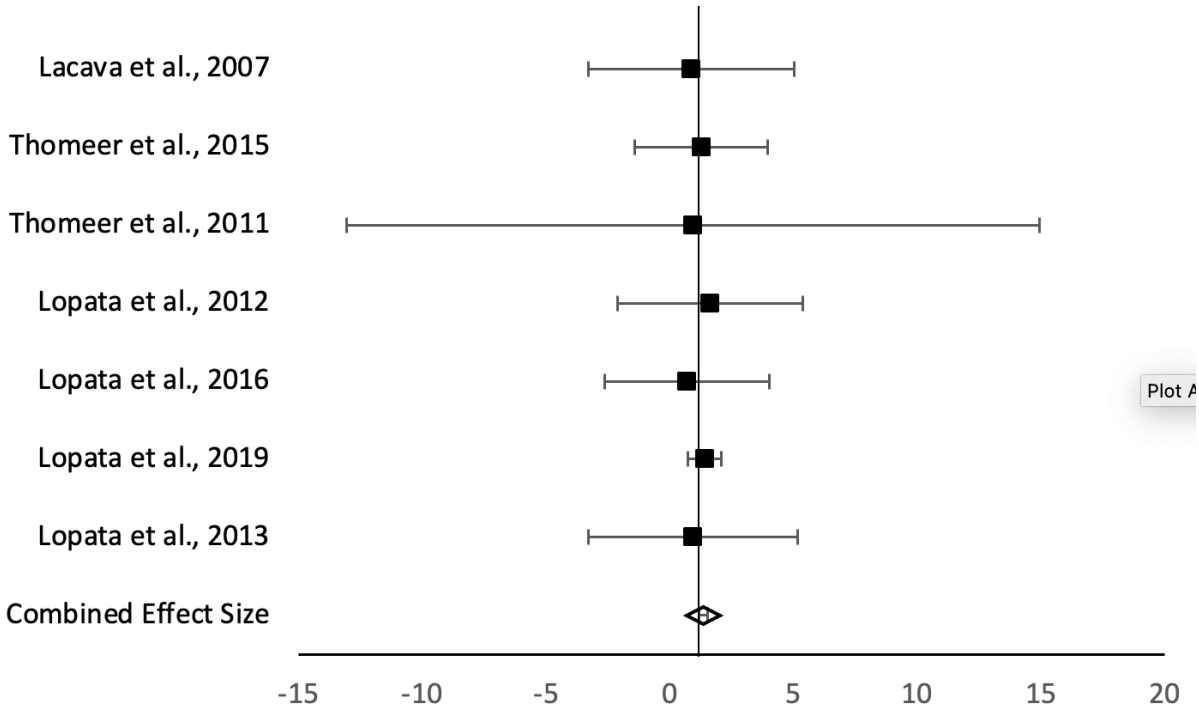
Authors: Davidson, et al.(2021) Study 2
Total
5
Average
1.67
Rating
Low

Appendix G

Omitting Davidson:
the meta-analysis revealed a significant and large treatment effect for Mind Reading on Emotion Identification d=1.39, 95%CI [0.92,1.86]). Results revealed no significant heterogeneity in effect size across all eight studies Q=2.36, I^2=0.0%.
Egger's regression test which was not significant (p=0.664).



Omitting Davidson and Weinger:
the meta-analysis revealed a significant and large treatment effect for Mind Reading on Emotion Identification d=1.35, 95%CI [1.15,1.56]). Results revealed no significant heterogeneity in effect size across all eight studies Q=0.35, I^2=0.0%.
Egger's regression test which was not significant (p=0.150).


Lacava et al., 2007
Thomeer et al., 2015
Thomeer et al., 2011
Lopata et al., 2012
Lopata et al., 2016
Lopata et al., 2019
Lopata et al., 2013
Combined Effect Size
-15
-10
-5
0
5
10
15
20

Appendix H

Including all ten studies:
the meta-analysis revealed a significant and large treatment effect for Mind Reading on Emotion Identification d=1.07, 95%CI [1.09,1.13]). Results revealed no significant heterogeneity in effect size across all eight studies Q=3.58, I^2=0.0%.
Egger's regression test which was not significant (p=0.144).