Case Study 1: An Evidence-Based Practice Review Report
Theme: Interventions for children with Special Educational Needs

HOW EFFECTIVE ARE COGNITIVE BEHAVIOURAL THERAPY- BASED SOCIAL SKILLS INTERVENTIONS FOR CHILDREN WITH HIGH FUNCTIONING AUTISM SPECTRUM DISORDERS (HFASD)?

SUMMARY
Severe and persistent impairment in social skills represents a core deficit in children with high functioning autism spectrum disorders (HFASD) and may contribute to poor academic outcomes, peer rejection and psychopathology (Little, 2001). Cognitive Behavioural Therapy (CBT) has been identified as one possible intervention to help children with HFASD change how they think, feel and respond in social situations. This review examines the effectiveness of CBT–based social skills interventions for school-age children with high functioning autism spectrum disorders (HFASD). Five selected studies were analysed using Kratochwill’s (2003) coding protocol and the weight of evidence was established according to Gough’s (2007) Weight of Evidence Framework. Preliminary evidence concludes that CBT is a promising intervention to treat social deficits in children with HFASD. However, the methodological limitations such as small sample size and lack of diverse population samples in the existing studies warrant further address. Recommendations are given for future research.
INTRODUCTION

The lack of appropriate social skills represents a pervasive deficit and lifelong struggle for children with high-functioning autism spectrum disorders (HFASD). These social deficits are primarily centred around social awareness, social cognition, social motivation, social reciprocity, social communication and social interaction skills (American Psychiatric Association, 2013; Attwood, 2004). Introspection and the ability to infer what others are thinking and feeling (i.e., “Theory of mind”) are particularly difficult for those with ASD (Frith & Happe, 1999). The weak central coherence theory of ASD (Frith, 1989) suggests that they have a limited ability to understand the context in which events occur, an inability to see the “big picture.” This results in individuals with autism taking things literally, such as not being able to understand metaphors, nuances or sarcasm. They also have a limited repertoire of behavioural responses to regulate emotional arousal (Attwood, 2007). Such difficulties may lead to poor social functioning and children with HFSAD are often reported to be socially isolated (Little, 2001).

Previous research on the efficacy of social skills training for children with ASD have received mixed reviews (Reichow & Volkmar, 2010; White, Koenig, & Scahill, 2007). Cognitive Behavioural Therapy (CBT) has been identified as one approach that can help children with HFASD learn self-awareness, self-control, and more constructive strategies to regulate emotions, as well as improve social cognition and social performance.

---

1 In the recent update of DSM-5 (American Psychiatric Association, 2013), Autism Spectrum Disorder (ASD) includes previous labels of autism disorder, Asperger’s Syndrome, and pervasive developmental disorder with the presence of co-occurring intellectual impairment listed as a specifier. Individuals with ASD and no co-occurring intellectual impairment are referred to as having High Functioning Autism Spectrum Disorder (HFASD). In line with the DSM-5 and for consistency purposes, this review will use the following terminology: HFASD (to refer to ASD with no intellectual disability and have at least average receptive language skills) and ASD (to refer to the broad spectrum, regardless of intellectual ability).
COGNITIVE BEHAVIOURAL THERAPY

CBT is broadly defined as a structured, time-bound, educational approach that focuses on developing cognitive and behavioural problem-solving strategies (Friedburg & McClure, 2002; Dobson & Dobson, 2009).

The cognitive aspects of CBT primarily focus on identifying and modifying maladaptive thoughts, attitudes and beliefs, as well as improving social understanding. In tandem with the emotional-behavioural aspects of CBT, the child with HFASD is encouraged to be more consciously attuned to his/her emotional state, changing the way he/she perceives social situations and responds socially, thereby increasing adaptive functioning (Craske, 2010). Craske (2010) has categorised the strategies and techniques used in CBT as follows:

(a) Cognitively based

Cognitively based strategies aim to challenge negative cognitions and replace them with more adaptive thoughts and beliefs. Modelling and cognitive rehearsals of appropriate strategies are useful teaching methods to help the child identify triggers that elicit specific cognitions, practice internal verbalisations (e.g. “self-talk”) and control their behaviour (Meichenbaum & Goodman, 1971). The use of concept clarification (e.g. “What is friendship?”) and Socratic questioning are also popular strategies in CBT to introduce novel social constructs and help the child gain insight into their maladaptive cognitions (Rotheram-Fuller & MacMullen, 2011).

(b) Skills and reinforcement based

Skills and reinforcement based strategies aim to develop the child’s problem-solving and coping skills. Problem solving skills include providing the child with schemas to perceive and analyse novel social situations or problems, identify the most appropriate strategy to resolve the problem, consider the consequences of...
actions, and evaluate the response outcome (Bauminger, 2002). Coping skills such as relaxation exercises, positive self-statements and self-monitoring of emotional states are introduced to help the child develop self-efficacy in emotion regulation and management of future stressors (Meichenbaum & Cameron, 1973). Through direct teaching and behavioural rehearsals such as role-play, the child can learn and practice the skills in a safe environment and immediate feedback is provided to reinforce the positive behaviour (Lopata et al., 2010). Affective education may also be introduced in this component to teach the child how to recognise emotions in self and others, as well as developing appropriate emotional responses (Attwood, 2004).

(c) Exposure based

Exposure based strategies aim to provide systematic, controlled and repeated exposure to difficult situations / stimuli such that the child no longer perceives the situation / stimuli negatively and/or avoids it (Craske, 2010).

The acquisition and performance of these three types of skills are the main goals of CBT-based social skills interventions, and emphasis is put on practising what has been learnt (i.e. homework) to promote sustainable change. The logical, practical, and structured approach of CBT makes it a suitable intervention for children with ASD who are not cognitively impaired, have sufficient language competency, and can benefit from addressing cognitive biases, deficits in affective knowledge, and social-behavioural performance.

PSYCHOLOGICAL UNDERPINNINGS

CBT is grounded in cognitive, behavioural, and learning principles of psychology (Dobson & Dobson, 2009).
The primary assumption of cognitive principles within CBT is that thoughts determine emotional regulation and behaviour. Therefore changing dysfunctional thoughts can lead to lowered emotional distress and improved social performance (Beck, 1972; Lazarus & Folkman, 1984).

The implementation of most CBT strategies is grounded in behavioural principles of psychology, particularly operant conditioning (Skinner, 1953). Operant conditioning principles dictate that the use of positive reinforcements, sometimes combined with punishers, can change behaviour based upon its consequences. For instance, receiving praise constitute positive reinforcement, which is a favourable consequence that increases the likelihood of a targeted behaviour occurring (e.g. volunteering responses in class), but punishers (e.g. being ignored by the teacher or laughed at by peers) which are aversive consequences will decrease the likelihood of the behaviour occurring again. Response cost procedures have been introduced in CBT interventions to reduce problem behaviours and foster skills acquisition (Lopata et al., 2008; 2010).

Bandura’s (1977) social learning theory helps to bridge both cognitive and behavioural principles within CBT. It postulates that people can learn from modelling and observing others’ behaviour, attitudes and outcomes of those behaviours. Two important components of learning – motivation and self-efficacy – are also important mediators of cognition, such as to set goals, think about actions and consequences, and engage in self-monitoring and evaluation (Bandura, 1973).

CBT-based social skills interventions thus encapsulate a reciprocal interaction between cognitive, behavioural and learning factors to help people make sense of complex situations (‘cognitive’), gain access to a repertoire of solutions (‘social learning’) and influence their reactions to situations (‘behaviour’).
RATIONALE

The relative strengths in cognitive and language abilities amongst children with HFASD tend to mask the manifestations of their social deficits. Unlike their lower cognitively functioning ASD peers, they are not flagged up so readily by parents and teachers, hence impeding their access for early intervention (Klin & Volkmar, 2000).

Lack of early intervention also suggests that such social deficits persist and become more severe in adolescence and adulthood where the nuances of social interactions increase in complexities (Billstedt, Gillberg & Gillberg, 2011)

Educational Psychologists (EPs) play an important role in ensuring that the needs of this group of children are not overlooked, and can consider CBT as a possible early intervention for children with HFASD to develop their social skills and improve their capacity for developing long term, quality social relationships.

There has been a paucity of research on how CBT can be applied to address social skills deficits in children with HFASD. Previous reviews have largely focused on either general social skills interventions with little differentiation between HFASD from the broad ASD population (Reichow & Volkmar, 2010; Rao, Beidel, & Murray, 2008; White, Koenig & Scahill, 2007) or the use of CBT to address issues of anxiety and/or core deficits (not specific to social skills) for children with ASD (Daniel & Wood, 2013; Rotheram-Fuller & MacMullen, 2011).

To the best of this reviewer’s knowledge, there are no similar reviews specific to the use of CBT as a social skills intervention for children with HFASD found in literature. Hence, this review aims to provide a critical appraisal of the current evidence base and answer the following question:
How effective are Cognitive Behavioural Therapy (CBT)-based social skills interventions for children with HFASD?

CRITICAL REVIEW OF THE EVIDENCE BASE

A comprehensive search was conducted in January 2014 to identify studies relevant to the research question. The search sources included PsychINFO, ERIC, Medline bibliographic databases, Google scholar search engine and ancestral searches of relevant articles. Figure 1 shows the search terms and combinations that were used and the stages of the systematic search.

Across the search sources, 45 studies were initially surfaced after duplications had been removed. The studies identified were screened using the inclusion / exclusion criteria shown in Appendix A. Based on their titles and abstracts, studies that did not meet the inclusion criteria were excluded. Full texts of 12 studies were obtained and checked against the inclusion criteria.
A summary of the five studies included in this review can be found in Appendix B. Four studies reviewed were conducted in the United States while one was conducted in Canada. The reasons for excluding the other 7 studies are provided in Appendix C.

The five studies were assessed using the Weight of Evidence Framework (Gough, 2007), as summarised in Tables 1 and 2 (see Appendix D for full description of the weighting of studies).

Table 1

Weight of Evidence Framework (Adapted from Gough, 2007)

<table>
<thead>
<tr>
<th>Weight of Evidence A</th>
<th>Weight of Evidence B</th>
<th>Weight of Evidence C</th>
<th>Weight of Evidence D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of methodology:</td>
<td>Relevance of Methodology:</td>
<td>Relevance of evidence to the review question:</td>
<td>Overall weight of evidence:</td>
</tr>
<tr>
<td>Methodological quality was assessed using the quality indicators outlined in Kratochwill's (2003) Literature Review Coding Protocol from the APA Task Force on Evidence Based Interventions in Schools Psychology (see Appendixes E1 – E5 for the coding protocols).</td>
<td>All studies used either experimental or quasi-experimental designs with pre and post measures for all outcomes related to social skills, which was judged to be appropriate for the review question.</td>
<td>The relevance of the study was assessed by the selection of primary intervention approach (i.e. use of CBT), participants (i.e. with a diagnosis of HFASD) and primary treatment setting.</td>
<td>One study was assessed to have ‘low’ weight of evidence while the other four studies have ‘medium’ weight of evidence.</td>
</tr>
</tbody>
</table>
Table 2

Weighting of Evidence

<table>
<thead>
<tr>
<th>Studies</th>
<th>(A) Quality of Methodology</th>
<th>(B) Relevance of Methodology</th>
<th>(C) Relevance of evidence to the review question</th>
<th>(D) Overall Weight of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Koning, Magill-Evans, Volden, &amp; Dick (2013)</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>DeRosier, Swick, Ornstein-Davis, Sturtz-McMillen &amp; Matthews (2011)</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>Wood et al. (2009)</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
</tr>
</tbody>
</table>

Aims and Social Targets of the Reviewed Studies

All five studies included behavioural training components to enhance social interaction skills (e.g. initiating social conversations and peer interaction). Socio-cognitive training aims included increasing general social awareness and knowledge (five studies); teaching emotion recognition, with or without emotion regulation (four studies); social motivation (three studies); Theory of Mind and perspective taking (two studies); and problem solving (one study). All five studies did not indicate specific targets for social awareness, social knowledge and emotion recognition, hence it was not possible to assess the range of skills and emotions taught.
Participants

All five studies included a series of inclusion/exclusion criteria to ensure that participants were appropriate for the intervention\(^2\). The total number of participants ranged from 15 (Koning et al., 2013) to 55 (DeRosier et al., 2011). As with any intervention study with special needs populations, particularly as the phenotype is highly specific, most prior studies examining social skills training for this population of children with HFASD have included very few participants – usually fewer than 20 participants (Rao et al., 2008). Based on previous literature, a medium effect size was estimated for the purposes of power-calculations. Cohen’s (1998) definition of medium (d<0.5) and an alpha-level of 0.05 were used to calculate required sample sizes. All five studies did not meet the Kratochwill (2003) criteria for sufficient sample size (i.e. minimum of 64 participants in between-group design for a medium effect size at an alpha-level of 0.05) and this affected their weightings of evidence (see Section A – ‘Analysis’ in Appendix D). This may be reflective of the recruitment difficulties often ascribed to HFASD populations. The small sample size in all the studies is an obvious limitation because it increases the risk of Type-1 errors and affects statistical power. Limited power could possibly explain the insignificant results on some of the outcome measures (e.g. Koning et al., 2013; Stichter et al., 2012).

With the exception of Stichter et al. (2012) which did not have a control group, the other four studies ensured that there were a similar number of participants in the intervention and control groups and matched for common characteristics (e.g. age, gender, socioeconomic backgrounds, receptive language scores).

\(^2\) In addition to meeting the diagnostic criteria of HFASD, participants are required to have at least average cognitive and language functioning, meet age requirements (i.e. between 6 to 12 years old) and show an absence of severe behavioural problems or comorbid diagnoses.
It is also noteworthy to point out the homogeneity of the samples in the five studies. Participants were primarily male, Caucasian, and from middle-class backgrounds. Of the 81 intervention participants across the five studies, only 4 were females. Although the prevalence of autism is higher amongst males than females (Rice, 2009), the lack of diversity amongst the participants limits the generalisability of the findings to the broader population of children with HFASD. These considerations affected the weightings of evidence for the ‘Relevance of Methodology’ (Section B) in Appendix D.

Agents and Setting

In all five studies, trained staff (psychology students, CBT therapists) provided the interventions, and the interventions took place at either private community-based practices (Koning et al., 2013; DeRosier et al., 2011) or at a university (Stichter et al., 2012; Lopata et al., 2010; Wood et al., 2009). Three studies ensured that there was counterbalancing of change agents and this improved their weightings of evidence (see Section A – ‘Comparison Group’ in Appendix D).

Assessments

The most frequently used parent-report questionnaire by all five studies was the Social Responsiveness Scale (SRS; Constantino & Gruber, 2005). The SRS has been identified to be a sensitive diagnostic screening tool which can provide an overall measure of the severity of social deficits and evaluate treatment changes in the ASD population (White al., 2012). Other parent-report questionnaires that were used included the Behaviour Assessment System for Children, Second Edition (BASC-2; Reynolds & Kamphaus, 2004) by Lopata et al. (2010); Vineland Adaptive
Behaviour Scales (Sparrow, Cicchetti & Balla, 2005) by Koning et al. (2013) and other researcher-designed questionnaires that are specific to the intervention, e.g. the Adapted Skillstreaming Checklist (Lopata et al., 2008).

Two studies sought teachers’ perspectives of the participants’ intervention progress by administering the SRS – Teacher’s Version (Stichter et al., 2012; Lopata et al., 2010). Child self-report was used in two of the studies. Lopata et al. (2010) developed the Skillstreaming Knowledge Assessment (SKA) to test the participants’ ability in identifying socially appropriate behaviour in a series of social situations. The SKA was reported by the authors to have an internal consistency reliability coefficient of .94 and an interrater reliability coefficient of .75. DeRosier et al. (2011) evaluated the participants’ perceived sense of self-efficacy using the Social Self-Efficacy Scale (Ollendick & Schmidt, 1987).

Some studies used social cognitive assessments to address specific skills targeted in the intervention, such as the Diagnostic Analysis of Nonverbal Accuracy 2 (DANVA2; Nowicki, 1997) which assesses participants’ ability to identify emotions through facial expressions (Stichter et al., 2012; Lopata et al., 2010). Other assessments that were used included the Child and Adolescent Social Perception Measure (CASP; Magill-Evans, Koning, Cameron-Sadava, & Manyk, 1995) by Koning et al. (2013), Reading the Mind in the Eyes Test (Baron-Cohen, Wheelwright, Raste, & Plumb, 2001), Theory of Mind test (Baron-Cohen, Leslie, & Frith, 1985) and Faux Pas Stories (Baron-Cohen, O’Riordan, Stone, Jones, & Plaisted, 1999) by Stichter et al. (2012).

For using measures with good psychometric characteristics (i.e. $\alpha \geq .85$) and multiple methods from at least two sources, three studies (Koning et al., 2013; DeRosier et al., 2011; Lopata et al., 2010) were awarded “high” weightings of
evidence for ‘Measures’ (see Section A in Appendix D). Although Stichter et al. (2012) used multiple assessment methods from multiple sources, some of its outcome measures did not report reliable psychometric scores (i.e. at least $\alpha \geq .70$), hence it was awarded “medium” weighting of evidence.

Wood et al. (2009) was not awarded any weighting because it had collected data using only one method from one source (i.e. parent-report questionnaire), thus did not meet criteria.

Outcomes

The social deficit in HFASD is multifaceted and the use of Lemerise and Arsenio’s (2000) integrated model of emotion processes and cognition in social information processing (see Figure 2) may be useful for framing and comparing the outcomes derived from the five studies (see Table 3).

<table>
<thead>
<tr>
<th>Stages</th>
<th>Type of Outcomes</th>
<th>Study</th>
<th>Outcome Measure Used</th>
<th>Effect Size / Descriptor</th>
<th>Overall Weight of Evidence</th>
</tr>
</thead>
</table>
Large | Medium |
|        |                   |       | Stichter, O’Connor, Herzog, Lieheimer & McGhee (2012) | Diagnostic Analysis of Non-Verbal Accuracy – 2, Child Facial Expressions | Cohen’s $d = 0.07$
Small | Low |
|        |                   |       |                       | Reading the Mind in Eyes Test | Cohen’s $d = 0.13$
Small | Medium |
|        |                   | • Social knowledge – recognising and understanding social norms and rules | Lopata, Thomeer, Volker, Toomey, Nida, Lee, Smerbeck & Rodgers (2010) | Diagnostic Analysis of Non-Verbal Accuracy – 2, Child Facial Expressions | Cohen’s $d = 0.532$
Medium | Medium |
|        |                   |       | Koning, Magill-Evans, Volden, & Dick (2013) | Social Knowledge Test | Partial eta squared = 0.68
Large | Medium |
Large | Medium |
|        | • Social awareness | Stichter, O’Connor, Herzog, Lieheimer & McGhee (2012) | Social Responsiveness Scale | Cohen’s $d = 0.33$
(parent measure)
Small | Low |
|        |                   |       |                       | Cohen’s $d = 0.31$
(teacher measure)
Small | Medium |
|        |                   |       | DeRosier, Swick, Ornstein-Davis, Sturtz-McMillen & Matthews (2011) | Social Responsiveness Scale | Cohen’s $d = 0.69$
Medium | Medium |
<p>|        |                   |       | Wood et al. (2009) | Social Responsiveness Scale | Not reported. Insufficient information provided to calculate effect size. | Medium |</p>
<table>
<thead>
<tr>
<th>Stages</th>
<th>Type of Outcomes</th>
<th>Study</th>
<th>Outcome Measure Used</th>
<th>Effect Size / Descriptor</th>
<th>Overall Weight of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.</td>
<td>• Interpreting Cues</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Inferring emotional state / affective cues (e.g. facial expression,</td>
<td>Koning, Magill-Evans, Volden, &amp; Dick (2013)</td>
<td>Child and Adolescent Social Perception Measure (Nonverbal Cues Score)</td>
<td>Partial eta squared = 0.45</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td>tone of voice, gestures) from peers’ nonverbal cues</td>
<td></td>
<td></td>
<td>Large</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Social cognition – decoding social cues in social contexts</td>
<td>Stichter, O’Connor, Herzog, Lieheimer &amp; McGhee (2012)</td>
<td>Social Responsiveness Scale</td>
<td>Cohen’s $d = 0.61$</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(parent measure)</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cohen’s $d = 0.27$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(teacher measure)</td>
<td>Small</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Understanding of idiomatic language</td>
<td>DeRosier, Swick, Ornstein-Davis, Sturtz-McMillen &amp; Matthews (2011)</td>
<td>Social Responsiveness Scale</td>
<td>Not reported as there</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>were no difference between means found</td>
<td>were no difference</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>between means found</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Not reported. Insufficient</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>information provided to calculate effect size.</td>
<td>information provided</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Theory of Mind – Perspective taking</td>
<td>Wood et al. (2009)</td>
<td>Social Responsiveness Scale</td>
<td>Not reported. Insufficient</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>information provided to calculate effect size.</td>
<td>information provided</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stages</td>
<td>Type of Outcomes</td>
<td>Study</td>
<td>Outcome Measure Used</td>
<td>Effect Size / Descriptor</td>
<td>Overall Weight of Evidence</td>
</tr>
<tr>
<td>--------</td>
<td>------------------</td>
<td>-------</td>
<td>----------------------</td>
<td>--------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>3.</td>
<td>Clarification of Goals</td>
<td>• Social motivation</td>
<td>Stichter, O’Connor, Herzog, Lieheimer &amp; McGhee (2012)</td>
<td>Social Responsiveness Scale</td>
<td>Cohen’s $d = 0.60$ (parent measure) Medium Cohen’s $d = 0.41$ (teacher measure) Small</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DeRosier, Swick, Ornstein-Davis, Sturtz-McMillen &amp; Matthews (2011)</td>
<td>Social Responsiveness Scale</td>
<td>Cohen’s $d = 0.67$ Medium</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Wood et al. (2009)</td>
<td>Social Responsiveness Scale</td>
<td>Not reported. Insufficient information provided to calculate effect size.</td>
</tr>
<tr>
<td>4.</td>
<td>Response Access or Construction</td>
<td>• Access possible behavioural / social responses in a specific social situation</td>
<td>Koning, Magill-Evans, Volden, &amp; Dick (2013)</td>
<td>Social Knowledge Test</td>
<td>Partial eta squared $= 0.68$ Large</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Metacognition</td>
<td>Stichter O’Connor, Herzog, Lieheimer &amp; McGhee (2012)</td>
<td>Behaviour Rating Inventory of Executive Function – Metacognition</td>
<td>Cohen’s $d = 0.42$ Small</td>
</tr>
<tr>
<td>5.</td>
<td>Response Decision</td>
<td>• Problem solving • Response evaluation • Understanding consequences of one’s actions • Appropriate social responding • Self-efficacy evaluation - to regulate emotions</td>
<td>Stichter, O’Connor, Herzog, Lieheimer &amp; McGhee (2012)</td>
<td>Behaviour Rating Inventory of Executive Function – Behavioural Regulation</td>
<td>Test of Problem Solving – 3</td>
</tr>
<tr>
<td>Stages</td>
<td>Type of Outcomes</td>
<td>Study</td>
<td>Outcome Measure Used</td>
<td>Effect Size / Descriptor</td>
<td>Overall Weight of Evidence</td>
</tr>
<tr>
<td>--------</td>
<td>------------------</td>
<td>-------</td>
<td>----------------------</td>
<td>--------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>6.</td>
<td>Behaviour Enactment</td>
<td>• Peer interaction</td>
<td>Koning, Magill-Evans, Volden, &amp; Dick (2013)</td>
<td>Peer Interaction Measure</td>
<td>Partial eta squared = 0.29 (Large)</td>
</tr>
</tbody>
</table>
|        |                   | • Sharing ideas  
• Demonstrating turn taking in conversations  
• Social communication | Stichter O’Connor, Herzog, Lieheimer & McGhee (2012) | Social Responsiveness Scale | Cohen’s $d = 0.74$ (parent measure)  
Medium               
Cohen’s $d = 0.40$ (teacher measure)  
Small             | Low |
|        |                   |       |                      |                          |                           |       |
|        |                   |       |                      |                          |                           |       |
|        |                   | Wood et al. (2009) | Social Responsiveness Scale | Not reported. Insufficient information provided to calculate effect size. | Medium |
|        |                   | DeRosier, Swick, Ornstein-Davis, Sturtz-McMillen & Matthews (2011) | Social Responsiveness Scale | Cohen’s $d = 0.94$ (Large) | Medium |
Large               
Cohen’s $d = 0.693$ (teacher measure)  
Medium             | Medium |
<p>| | | | | | | |
|        |                   |       |                      |                          |                           |       |
|        |                   |       |                      |                          |                           |       |
|        |                   |       |                      |                          |                           |       |
|        |                   |       |                      |                          |                           |       |</p>
<table>
<thead>
<tr>
<th>Stages</th>
<th>Type of Outcomes</th>
<th>Study</th>
<th>Outcome Measure Used</th>
<th>Effect Size / Descriptor</th>
<th>Overall Weight of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Socialisation</td>
<td>Overall socialisation skills</td>
<td>Koning, Magill-Evans, Volden, &amp; Dick (2013)</td>
<td>Social Responsiveness Scale</td>
<td>Partial eta squared = 0.00 Small</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Vineland Adaptive Behaviour Scales- II (Socialization scale)</td>
<td>Partial eta squared = 0.01 Small</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stichter, O’Connor, Herzog, Lieheimer &amp; McGhee (2012)</td>
<td>Social Responsiveness Scale</td>
<td>Cohen’s $d = 0.75$ (parent measure) Medium</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cohen’s $d = 0.39$ (teacher measure) Small</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>DeRosier, Swick, Ornstein-Davis, Sturtz-McMillen &amp; Matthews (2011)</td>
<td>Achieved Learning Questionnaire</td>
<td>Cohen’s $d = 0.75$ Medium</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Social Self-Efficacy Scale Not reported as there were no difference between means found</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lopata, Thomeer, Volker, Toomey, Nida, Lee, Smerbeck &amp; Rodgers (2010)</td>
<td>Social Responsiveness Scale</td>
<td>Cohen’s $d = 0.625$ (parent measure) Medium</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cohen’s $d = 0.711$ (teacher measure) Medium</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Adapted Skillstreaming Checklist</td>
<td>Cohen’s $d = 0.584$ (parent measure) Medium</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cohen’s $d = 1.421$ (teacher measure) Large</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wood et al. (2009)</td>
<td>Social Responsiveness Scale</td>
<td>Cohen’s $d = 0.77$ Medium</td>
<td>Medium</td>
</tr>
</tbody>
</table>

**Note**
- Partial eta squared values are considered small at .01, medium at .09 and large at .025 (Bakeman, 2005).
- Cohen’s $d$ values are considered small at 0.2, medium at 0.5 and large at 0.8 (Cohen, 1988).
The model describes the interplay between elements of social context, social perception, social cognition, social behaviour, social interaction, as well as motivation and emotional processes. Within the model, an interaction is composed of the following key stages:

**Stage 1 – Encoding of social and emotional cues**

All five studies stated outcomes that will enable participants to be more effective in encoding social and emotional cues. Three studies looked at ‘emotion recognition’ and achieved varied effects. While Koning et al. (2013) reported a large effect size, the other two studies reported small (Stichter et al., 2012) and medium effect sizes (Lopata et al., 2010) respectively. Although the three studies demonstrated some improvements in emotion recognition, caution needs to be exercised in the interpretation of the results as the treatment effects were not well-clarified in terms of the type and complexity of emotions attained by the participants.

Similarly, the outcome of general ‘social awareness’ achieved small to medium effect sizes based on two studies with a ‘low’ and ‘medium’ overall weight of evidence (Stichter et al., 2012; DeRosier et al., 2011). Although Wood et al. (2009) looked at ‘social awareness’ as an outcome too, it did not report or provide sufficient data to calculate an effect size.

Although Koning et al. (2013) and Lopata et al. (2010) reported large effect sizes for the ‘recognition of social norms and rules’, both studies only attained ‘medium’ overall weight of evidence. These findings thus suggest that the acquisition of social knowledge is at best a modest outcome within this stage.
**Stage 2 - Interpretation of social and emotional cues**

Only one study which had a ‘medium’ overall weight of evidence looked at ‘inference of peers’ emotional and affective cues’ and reported a large effect size (Koning et al., 2013). Three studies focused on ‘social cognition’ (i.e. decoding cues in the social context) but only one study (Stichter et al., 2012) reported a small (teacher’s report) to medium (parent’s report) effect size. The other two studies did not report or provide sufficient data to calculate an effect size (DeRosier et al, 2011; Wood et al., 2009).

Lopata et al. (2010) posited that the teaching of idiomatic language will enable participants to be more competent in the interpretation of non-literal language, but found a small effect size. Similarly, Stichter et al. (2012) reported significant improvements in participants’ ability to recognise the social faux pas portrayed in given vignettes but pre-post measures only reported a small effect size. Taking into consideration that Stichter et al. (2012) received a ‘low’ overall weight of evidence primarily due to its low quality and relevance of methodology, it cannot be postulated with confidence that the intervention will enable participants to recognise social faux better.

**Stage 3 – Clarification of social goals are clarified**

Findings related to this stage needs to be interpreted with caution as none of the studies observed the participants in their natural social environments to provide an objective assessment of the participants’ social goals (e.g. avoidant, hostile or cooperative goal) within a social context.

Three studies of ‘low’ to ‘medium’ overall quality that looked into ‘social motivation’ relied only on parent-report questionnaires. They reported a medium
effect size for this outcome (Stichter et al., 2012; DeRosier et al., 2011). Wood et al. (2009) did not report or provide sufficient data to calculate an effect size. As the parent informants in these three studies were either heavily involved in the intervention and/or knew about the treatment goals and procedures, their perceptions of their children’s progress may be subjected to bias and should be further verified through more objective direct child observations.

Stage 4 – Response access or construction

Two studies evaluated different outcomes in this stage. Koning et al. (2013) focused on teaching possible solutions to deal with difficult situations and reported a large effect size. Although Koning et al. (2013) had a 'medium' overall weight of evidence, its relevance of evidence to the review question was rated 'low' due to its primary treatment setting being in a private community-based practice. Hence, it may be difficult to replicate the findings in this study in EPs’ practice for children with HFASD in mainstream schools. Stichter et al. (2012) reported a small effect size for acquiring metacognitive strategies (e.g. initiation, planning, organising of thoughts). However, Stichter et al. (2012) was rated 'low' overall quality, suggesting that its small effect size should be carefully considered when evaluating the effectiveness of CBT in the development of metacognition.

Stage 5 – Response decision

Surprisingly, for CBT-based interventions expected to adopt a more problem-solving approach, only one study (Stichter et al., 2012) had evaluated ‘behavioural regulation’ and ‘problem solving’ outcomes that are imperative for this stage. Although Stichter et al. (2012) reported statistically significant findings for both
outcomes at the p < .05 level, the results should be treated with caution due to their small effect sizes, and taking into consideration that the study received ‘low’ weightings for quality and relevance of methodology, as well as overall quality.

**Stage 6 – Behaviour enactment**

All five studies evaluated similar outcomes in this stage. Koning et al. (2013) evaluated on participants’ effectiveness in peer interaction. Three studies evaluated on social communication and turn-taking skills (Stichter et al., 2012; DeRosier et al., 2011; and Wood et al., 2009) and one study evaluated on the participants’ social response (i.e. reduced tendency to escape from or avoid social contact) that will help to facilitate positive social interaction. The findings for this stage looks most promising as all the studies reported medium to large effect sizes on parent-reported measures, and four out of the five studies were awarded ‘medium’ overall weight of evidence.

**Overall socialisation skills**

All five studies utilised parent and/or teacher-reported measures to evaluate overall socialisation skills. Four out of the five studies reported medium effect sizes. One study (Koning et al., 2013) reported a small effect size but this could possibly be attributed to its small sample size. There was some slight discrepancy noted between parent and teacher ratings in this outcome (Stichter et al., 2012; Lopata et al., 2010) and this may suggest that gains at the treatment site have not been fully generalised to some extent in other social settings such as in school.
CONCLUSION

The five studies reviewed examined the effectiveness of CBT-based social skills intervention for children with HFASD. Four studies have ‘medium’ weighting of evidence and one study (Stichter et al., 2012) has ‘low’ weighting of evidence due to its low quality and relevance of methodology.

Overall, there is preliminary evidence that CBT is a promising intervention to treat social deficits in children with HFASD and should be included in the “tool box” of EPs. However, caution should be exercised in replicating the studies, taking into consideration methodological limitations in existing literature.

CBT’s strong emphasis on structure, use of short discrete activities, concrete socio-cognitive and behavioural strategies, as well as, performance feedback makes it a very suitable intervention for children with HFASD. Often, children with HFASD tend to overcompensate by using their logical thinking to evaluate social situations and lack emotion regulation skills. The integrated emphasis of CBT, combined with the adapted techniques for HFASD, can thus help them to learn new ways to understand and monitor their thoughts, feelings and actions.

The five studies’ pre and post measures indicate that all participants have made some progress in one or more social skills domains, thereby highlighting the strength of the intervention. Based on the mapping of outcomes to Lemerise and Arsenio’s (2000) model, it also appears that all five studies have placed most emphasis on Stages 1 (encoding of cues) and 6 (behaviour enactment), followed by Stages 2 (interpretation of cues) and 3 (clarification of goals). Less attention has been paid to Stages 4 (response access and construction) and 5 (response decision). As the range of targeted social skills differ across the five studies, it is unclear at present point if the use of CBT is more effective in a specific social skill
domain (e.g. social awareness) or similar across all domains. Selection of optimal social skill targets to be incorporated in the CBT-based interventions warrant further investigation.

One limitation across the five studies involved the heavy reliance of parent-reported measures completed by parents who are aware of the participants’ treatment condition and intervention procedures. While the significant improvements were also substantiated by child-reported measures, the threat of rater bias remains.

There is little evidence on generalisation and maintenance effects. All five studies were limited by small sample size and homogeneity in sample characteristics. Although the interventions have been shown to be effective in clinical settings, none of the studies looked into generalisation of behaviour gains in naturalistic settings (e.g. home and school). Four out of five studies also did not conduct any follow-up assessments post-treatment. Only one study (Stichter et al., 2009) demonstrated treatment gains were maintained at 3-month follow-up. While the data is encouraging, conclusions on longer term maintenance effects cannot be made.

**RECOMMENDATIONS**

Based on this review, several recommendations can be made for the use of CBT-based social skills intervention and for future research:

a) The design and implementation of a CBT-based intervention should consider the developmental needs of children with HFASD as they display a wide range of cognitive maturity and abilities. The CBT approach can be modified, for instance, incorporating lessons into play routines so that the content is easily accessible for younger children, providing concrete examples of social
constructs and involving parents or main caregivers in the treatment process so that they can complement support at home.

b) As social deficits are multidimensional (Kasari et al., 2001), Lemerise and Arsenio’s (2000) integrated model may be a useful framework for understanding socio-emotional information processing difficulties, planning and implementing intervention, and measuring changes in social skills for children with HFASD.

c) Treatment contexts must be considered. Given that children with HFASD may find it challenging to generalise skills to different contexts, as evident from the discrepancy of findings between parents and teachers in two of the studies, it appears imperative to consider schools as a primary context for CBT-based interventions to take place. In addition, as the role of the peer group becomes more important in late childhood and adolescence, it may be helpful to incorporate peer support in the CBT intervention to create more authentic social experiences.

d) Future research needs to tap on a more diverse population and have a larger sample size, so as to enhance the generalizability of the findings.

e) The use of multiple measures from a wide variety of sources is crucial for an objective evaluation of participants’ social-interaction skills. For instance, including direct child observations of operationally defined social behaviours children during peer interaction activities will provide a more robust picture of treatment effects. Observer bias should also be addressed (e.g. use of ‘blind’ observers).

f) Future research should attempt to demonstrate that CBT-based interventions are at least equivalent to another intervention (not ‘control/waitlist group’ or
‘no intervention), as well as, including longer-term follow-up measures to provide stronger evidence of efficacy and maintenance of treatment gains post-intervention.
REFERENCES


## Inclusion / Exclusion Criteria

<table>
<thead>
<tr>
<th></th>
<th>Inclusion Criteria</th>
<th>Exclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td><strong>Type of Publication</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The study is published in a peer-reviewed journal. This ensures methodological</td>
<td>The study is not published in a peer-reviewed journal (e.g. conference papers).</td>
</tr>
<tr>
<td></td>
<td>rigour as peer-reviewed journals have been assessed using similarly stringent</td>
<td></td>
</tr>
<tr>
<td></td>
<td>criteria.</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td><strong>Language</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The study is written in English, due to unavailability of resources for</td>
<td>The study is not written in English.</td>
</tr>
<tr>
<td></td>
<td>translation.</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td><strong>Research Design</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The study is an empirical examination of the effectiveness of CBT-oriented social</td>
<td>The study is not empirical in nature and does not look at the effectiveness of the</td>
</tr>
<tr>
<td></td>
<td>interventions.</td>
<td>intervention.</td>
</tr>
<tr>
<td></td>
<td>The study uses a casual research design (i.e. randomised controlled trials, group</td>
<td>The study does not have a control group and/or pre/post data measures.</td>
</tr>
<tr>
<td></td>
<td>comparison studies, and includes pre/post data).</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td><strong>Participants</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>School aged children (ages 6 to 12 years) with a confirmatory diagnosis of high</td>
<td>School aged children without a confirmatory diagnosis of HFASD; have below average</td>
</tr>
<tr>
<td></td>
<td>functioning autism spectrum disorder / Asperger’s syndrome / PDD-NOS, with average</td>
<td>receptive language skills and IQ.</td>
</tr>
<tr>
<td></td>
<td>receptive language skills and IQ.</td>
<td>Adolescents with HFASD.</td>
</tr>
<tr>
<td>5.</td>
<td><strong>Intervention</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The main intervention being evaluated is a CBT-based approach to target social</td>
<td>General social skills training based on the principles of applied behaviour</td>
</tr>
<tr>
<td></td>
<td>skills deficit. CBT techniques such as modelling, cognitive rehearsals, affective</td>
<td>analysis.</td>
</tr>
<tr>
<td></td>
<td>education, behavioural role-play, as well as provision of feedback and</td>
<td>Multi-component interventions in which CBT is not the primary treatment strategy</td>
</tr>
<tr>
<td></td>
<td>reinforcements should be used as the primary treatment strategy.</td>
<td>are excluded due to difficulty attributing the intervention effects to the use of</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CBT.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The CBT-based approach is used to treat anxiety and</td>
</tr>
</tbody>
</table>
not purely targeting social skills deficit.

| 6. Date | The study is published post 2008, based on the last social skills interventions for individuals with autism review by Reichow & Volkmar (2010) which accessed publication between 2001 to 2008. | The study is published in 2008 or earlier. |
## Summary of Reviewed Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Agents / Setting</th>
<th>Techniques</th>
<th>Outcomes</th>
</tr>
</thead>
</table>
| Koning, Magill-Evans, Volden & Dick (2013) | Male, ages 10 – 12 years | 2hr weekly group intervention led by 2 trained staff for 15 sessions | • Group intervention  
• Didactic instruction (concept clarification and social learning)  
• Active practice through:  
  - Cooperative play  
  - Role playing  
  - Modelling  
  - Games  
  - “out loud self-talk” opportunities for cognitive rehearsal  
  - Lego building activities  
  - Homework  
• Performance feedback through prompts/ praise and self-appraisal in one-to-one and group settings | • Significant main effects EX>CO on all children’s measures:  
  - Child and Adolescent Social Perception Measure: nonverbal cue score and emotion score  
  - Social Knowledge questionnaire  
  - Peer Interaction Measure to assess how participants generalise their social skills in a structured by naturalistic social situation  
• No significant interaction effects on all parent report measures:  
  - Vineland-II, Socialisation scale  
  - Social Responsiveness Scale |
<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Agents / Setting</th>
<th>Techniques</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stichter, O’Connor, Herzog, Lieheimer &amp; McGhee (2012)</td>
<td>Children ages 6 – 10 years</td>
<td>SCI-E adapted to elementary-age children with HFASD: • Affective education  • Structured, scaffolding that reviews a previously learned skill, and introducing a new skill in an instructional and hands-on structured learning format • Skill modelling • Opportunities to practice the skill in a structured activities • Closing activity or review</td>
<td>• Significant improvements Pre &lt; Post on:  Social Abilities • Social Responsiveness Scale (Parent) – all subscales • Social Responsiveness Scale (Teacher) – total; communication; motivation; reduced autistic mannerisms Theory of Mind • ToM: recognising social mistakes (faux pas test) Executive Functioning • Test of Problem Solving 3 (TOPS-3): sequencing the problem scenarios • Behavioural Rating Inventory of Executive Function (BRIEF) [parent reported measure]: metacognition and behavioural regulation</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Participants</td>
<td>Agents / Setting</td>
<td>Techniques</td>
<td>Outcomes</td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td>-----------------</td>
<td>------------</td>
<td>----------</td>
</tr>
</tbody>
</table>
| DeRosier, Swick, Ornstein-Davis, Sturtz-McMillen & Matthews (2011) | Children ages 8 – 12 years | 1hr weekly group intervention for HFASD conducted over 15 weeks by 2 trained staff | • Group intervention  
• Didactic instruction (concept clarification and social learning)  
• Parent training  
• Active practice through  
- Role playing  
- Modelling  
- Homework  
- Community-based activities | • Significant main effects EX > CO on all parents’ measures:  
- Social Responsiveness Scale (SRS): awareness, communication, motivation, reduced autistic mannerisms  
- Social Skills Scale – the Achieved Learning Questionnaire (ALQ)  
- Social self-efficacy  
• EX = CO on child’s reported loneliness scale and on self-reported social efficacy |
| | • EX: n = 27; [26 males; 1 female]; training adapted to HFASD (CA: 10.2)  
• CO: n = 28; [all males]; traditional social skills training, non-HFASD-specific (CA: 9.9) | Parent coached by community social missions |  |  |
| | Main treatment: private, community-based practice | Generalisation: community social tasks coached by parents |  |  |

DeRosier, Swick, Ornstein-Davis, Sturtz-McMillen & Matthews (2011) studied children ages 8 – 12 years with HFASD. The study involved 27 females and 26 males, with an average age of 10.2 years. The intervention consisted of a 1-hour weekly group intervention conducted over 15 weeks by 2 trained staff. The group intervention included didactic instruction, parent training, and active practice through role playing, modelling, homework, and community-based activities. The study found significant main effects EX > CO on all parents’ measures, including Social Responsiveness Scale (SRS) for awareness, communication, motivation, and reduced autistic mannerisms, Social Skills Scale – the Achieved Learning Questionnaire (ALQ), and social self-efficacy. The study concluded that EX = CO on child’s reported loneliness scale and on self-reported social efficacy.
<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Agents / Setting</th>
<th>Techniques</th>
<th>Outcomes</th>
</tr>
</thead>
</table>
| Lopata, Thomeer, Volker, Toomey, Nida, Lee, Smerbeck & Rodgers (2010) | Children ages 7 – 12 years          | 5-week group intervention, 5 days per week, 5 treatment cycles per day, led by professionals (students in education and psychology) | • Direct instruction  
• Modelling  
• Role-playing  
• Performance feedback  
• Transfer of learning  
• Affective education  
• Problem solving  
• Concept clarification of nonliteral language  
• Cooperative dyadic group activities  
• Homework  
• Response-cost behaviour system to increase and maintain pro-social behaviours and decrease problem behaviours  
• Parent training | Significant main effects EX>WL on parent rating measures:  
- Social Responsiveness Scale (SRS)  
- Adapted Skillstreaming Checklist (ASC) assessing skills taught in the social program  
- Behaviour Assessment System for Children, 2nd Ed (BASC-2) assessing withdrawal  

Significant main effects EX>WL on children’s measures:  
- Skillstreaming Knowledge Assessment (SKA) of skills taught in the social program  
- Comprehensive Assessment of Spoken Language (CASL): idiomatic language |

Each of the 5 daily 70 minute treatment cycles included:  
- 20 minute intensive instruction  
- 50 minute therapeutic activity  

Summer program on college campus using group rooms
<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Agents / Setting</th>
<th>Techniques</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wood et al.</td>
<td>Children ages 7 – 11 years</td>
<td>1.5hr weekly family CBT intervention conducted by therapists over 16 weeks:</td>
<td>• Vivo exposure supported by parent training and school consultation to promote social communication and emotion regulation skills</td>
<td>• Significant main effects EX&gt;WL on parent reported measures:</td>
</tr>
<tr>
<td>(2009)</td>
<td>• Total N = 19 (16 males) (CA: 9.37)</td>
<td>- 30 min with the child</td>
<td>• Social coaching</td>
<td>- Social Responsiveness Scale (SRS): Social Communication; Social Motivation; Social Awareness; and Social Cognition</td>
</tr>
<tr>
<td></td>
<td>• EX: n = 9</td>
<td>- 60 min with the parents/ family, implementing a version of the Building Confidence CBT program</td>
<td>• Guided conversations — use of Socratic questioning</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• WL: n = 10</td>
<td></td>
<td>• Peer Buddying and mentoring programmes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Behavioural reward system</td>
<td></td>
</tr>
</tbody>
</table>

Main treatment: university

Also provided on-site social coaching and consultation in school and at home

Note: EX = Experimental group; WL = Waitlist group; CO = Control group; CA = Chronological Age
## APPENDIX C

### Studies Excluded from the Review

<table>
<thead>
<tr>
<th>Study</th>
<th>Reasons for Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Cotugno, A. J. (2009). Social competence and social skills training and intervention for children with autism spectrum disorders. <em>Journal of Autism and Developmental Disorders, 39</em>(9), 1268-1277.</td>
<td>The study used a multidimensional structure and skill based approach to target social competence. CBT was not used as a primary treatment strategy. <em>(Criterion 5)</em></td>
</tr>
<tr>
<td>6 Solomon, M., Goodlin-Jones, B., &amp; Anders, T. (2004). A social adjustment enhancement intervention for high functioning autism, Asperger’s syndrome, and pervasive CBT was not used as a primary treatment strategy for social difficulties. <em>(Criterion 5)</em></td>
<td></td>
</tr>
<tr>
<td>Developmental Disorder NOS.</td>
<td>The study was published in 2004. (Criterion 6)</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------------------------------------------</td>
</tr>
</tbody>
</table>
Appendix D

Weighting of Studies

The five studies were assessed using the Weight of Evidence Framework (Gough, 2007).

A: Methodological Quality

Using the Kratochwill (2003) coding protocol, the studies were weighted on their ‘Measures’, ‘Comparison Group’ and ‘Analysis’.

**Measures**

<table>
<thead>
<tr>
<th>Weighting</th>
<th>Description</th>
</tr>
</thead>
</table>
| **High**  | • Used measures that produce scores whose internal consistency reliability is at least 0.85 for the majority of primary outcomes under investigation and for the current population under study.  
• Data collected using multiple (i.e. at least two assessment methods or approaches were used) methods (e.g. observational, self-reports, parent ratings) and from multiple (i.e. at least two) sources (e.g. parents, teachers, self).  
• Reported validity for all measures used and/or used a well-referenced, standardised or norm-referenced measure.  
• In the event that multiple primary outcome measures are used, the above criteria must be met for all primary outcome measures. |
| **Medium** | • Used measures that produce scores whose internal consistency reliability is at least 0.70 for the majority of primary outcomes under investigation and/or for the current population under study.  
• Data collected using either multiple methods and/or from multiple sources.  
• Reported validity for some measures used and/or used a well-referenced, standardised or norm-referenced measure.  
• In the event that multiple primary outcome measures are used, the above criteria must be met for at least 75% of the primary outcome measures. |
| **Low**    | • Used measures that produce somewhat reliable scores whose internal consistency reliability is at least 0.50 for the primary outcomes under investigation and/or for the current population under study.  
• Data collected using either multiple methods and/or from multiple sources.  
• A case for validity does not need to be presented.  
• In the event that multiple primary outcome measures are |
used, the above criteria must be met for at least 50% of the primary outcome measures.

### Comparison Group

<table>
<thead>
<tr>
<th>Weighting</th>
<th>Description</th>
</tr>
</thead>
</table>
| **High**  | - Used at least one type of ‘active’ comparison group (e.g. typical intervention, attention placebo, intervention element placebo, alternative intervention, pharmacotherapy)  
- Established initial group equivalency, preferably through random assignment of participants to intervention conditions  
- Demonstrated evidence of counterbalancing  
- Demonstrated low attrition or equivalent mortality at the post test.  
- Used at least a ‘no intervention group’ type of comparison (e.g. no intervention, wait list / delayed intervention, minimal contact).  
- Demonstrated evidence for at least two of the following: (a) group equivalence established; (b) counterbalancing, or (c) equivalent mortality with low attrition.  
- Conducted an intent-to-intervene analysis if equivalent mortality was not demonstrated. |
| **Medium**| - Used a comparison group.  
- Demonstrated evidence for at least one of the following: (a) group equivalence established; (b) counterbalancing, or (c) equivalent mortality with low attrition.  
- Conducted an intent-to-intervene analysis if equivalent mortality is not demonstrated. |
| **Low**   | - Used a comparison group.  
- Demonstrated evidence for at least one of the following: (a) group equivalence established; (b) counterbalancing, or (c) equivalent mortality with low attrition.  
- Conducted an intent-to-intervene analysis if equivalent mortality is not demonstrated. |
## Analysis

<table>
<thead>
<tr>
<th>Weighting</th>
<th>Description</th>
</tr>
</thead>
</table>
| **High**  | - Conducted an appropriate statistical analysis including: appropriate unit of analysis, familywise/experimentwise error rate controlled (if applicable) and a sufficiently large N.  
- Reported effect sizes or provided enough information for effect sizes to be calculated for all the primary outcomes.  
- Demonstrate significant primary outcomes for at least 75% of the total primary outcome measures.  

Demonstrated at least 3 of the following:  
- Conducted an appropriate statistical analysis including: appropriate unit of analysis and/or familywise/experimentwise error rate controlled (if applicable).  
- Sufficiently large N.  
- Reported effect sizes or provided enough information for effect sizes to be calculated for some of the primary outcomes.  
- Demonstrate significant primary outcomes for at least 50% to 74% of the total primary outcome measures (i.e. at least two out of four outcomes measures show statistically significant change in the desired direction for a given primary outcome construct). |
| **Medium** | Demonstrated at least 2 of the following:  
- Conducted an appropriate statistical analysis including: appropriate unit of analysis and/or familywise/experimentwise error rate controlled (if applicable)  
- Reported effect sizes or provided enough information for effect sizes to be calculated for some of the primary outcomes.  
- Demonstrate significant primary outcomes for at least 25% to 49% of the total primary outcome measures (i.e. at least one out of four outcomes measures show statistically significant change in the desired direction for a given primary outcome construct). |
| **Low**    |             |
Overall Methodological Quality

The overall methodological quality of the studies was calculated by assigning scores of:

- ‘3’ for ‘High’ weighting
- ‘2’ for ‘Medium’ weighting
- ‘1’ for ‘Low’ weighting
- ‘0’ for studies which did not meet the criteria for a ‘Low’ weighting

Scores were then averaged.

<table>
<thead>
<tr>
<th>Overall Methodological Quality</th>
<th>Average Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>At least 2.5</td>
</tr>
<tr>
<td>Medium</td>
<td>Between 1.5 and 2.4</td>
</tr>
<tr>
<td>Low</td>
<td>Less than 1.4</td>
</tr>
</tbody>
</table>

The overall methodological quality of the five studies is as follows:

<table>
<thead>
<tr>
<th>Studies</th>
<th>Measures</th>
<th>Comparison Group</th>
<th>Analysis</th>
<th>Overall Quality of Methodology</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Koning, Magill-Evans, Volden, &amp; Dick (2013)</em></td>
<td>Weighting: High</td>
<td></td>
<td></td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td>Score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2.3</td>
</tr>
<tr>
<td><em>Stichter, O’Connor, Herzog, Lieheimer &amp; McGhee (2012)</em></td>
<td>Weighting: Medium</td>
<td>Did not meet criteria</td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>1.33</td>
</tr>
<tr>
<td><em>DeRosier, Swick, Ornstein-Davis, Sturtz-McMillen &amp; Matthews (2011)</em></td>
<td>Weighting: High</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2.67</td>
</tr>
<tr>
<td><em>Lopata, Thomeer, Volker, Toomey, Nida, Lee, Smerbeck &amp; Rodgers (2010)</em></td>
<td>Weighting: High</td>
<td>Medium</td>
<td></td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td>Score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2.33</td>
</tr>
<tr>
<td><em>Wood et al. (2009)</em></td>
<td>Weighting: Did not meet criteria</td>
<td></td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1.33</td>
</tr>
</tbody>
</table>
B: Relevance of Methodology

This weighting reviews the appropriateness of the methodology for answering the review question: “How effective are cognitive behavioural therapy-based social skills interventions for children with high functioning autism spectrum disorders?”

<table>
<thead>
<tr>
<th>Weighting</th>
<th>Description</th>
</tr>
</thead>
</table>
| **High**  | • Used random assignment of participants.  
             • Used an ‘active’ comparison group (e.g. typical intervention, attention placebo, intervention element placebo, alternative intervention, and pharmacotherapy).  
             • Demonstrated group equivalences.  
             • Obtained pre and post measures for both intervention and comparisons groups. |
| **Medium**| • Used a comparison group.  
             • Demonstrated group equivalences.  
             • Obtained pre and post measures for both intervention and comparisons groups. |
| **Low**   | • Obtained pre and post measures for all groups involved. |

C: Relevance of Evidence to Review Question

This weighting is a review-specific judgement about the relevance of the focus of the evidence for the review question.

<table>
<thead>
<tr>
<th>Weighting</th>
<th>Description</th>
</tr>
</thead>
</table>
| **High**  | • Intervention is explicitly stated as a cognitive-behavioural therapy approach and used cognitive-behavioural therapy techniques as a primary treatment strategy for social skills difficulties.  
             • Used school-age children (ages 6 to 12 years) with a confirmatory diagnosis of high functioning autism spectrum disorders / Asperger’s Syndrome / Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS), with average receptive language and IQ.  
             • Primary treatment setting is in a mainstream school (i.e. more applicable as a potential intervention for EP practice)  
             • Measured social skills pre and post intervention. |
| **Medium**| • Intervention is not explicitly stated as cognitive-behavioural therapy approach but has incorporated cognitive-behavioural therapy techniques as a primary treatment strategy for social skills difficulties.  
             • Used school-age children (ages 6 to 12 years) with a confirmatory diagnosis of high functioning autism spectrum disorders / Asperger’s Syndrome / Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS), with average
• Receptive language and IQ.
• Primary treatment setting is in a classroom within an academic setting (e.g. university).
• Measured social skills pre and post intervention.

• Intervention does not incorporate cognitive-behavioural therapy techniques as a primary treatment strategy for social skills difficulties.
• Did not use school-age children (ages 6 to 12 years) with a confirmatory diagnosis of high functioning autism spectrum disorders / Asperger’s Syndrome / Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS), with average receptive language and IQ.
• Primary treatment setting is in a private / community-based practice.

D: Overall Weight of Evidence
This is an overall assessment of the extent to which the study contributes evidence to answer the review question. The weighting of evidence awarded for each score was as follows:

- ‘3’ for ‘High’ weighting
- ‘2’ for ‘Medium’ weighting
- ‘1’ for ‘Low’ weighting

These scores were then averaged to find the overall weight of evidence score.

<table>
<thead>
<tr>
<th>Overall Methodological Quality</th>
<th>Average Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High</strong></td>
<td>At least 2.5</td>
</tr>
<tr>
<td><strong>Medium</strong></td>
<td>Between 1.5 and 2.4</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>Less than 1.4</td>
</tr>
</tbody>
</table>
The overall weight of evidence for the five studies is as follows:

<table>
<thead>
<tr>
<th>Studies</th>
<th>(A) Quality of Methodology</th>
<th>(B) Relevance of Methodology</th>
<th>(C) Relevance of evidence to the review question</th>
<th>(D) Overall Weight of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Koning, Magill-Evans, Volden, &amp; Dick (2013)</td>
<td>Weighting Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td>Score 2</td>
<td>2</td>
<td>1</td>
<td>1.67</td>
</tr>
<tr>
<td></td>
<td>Score 1</td>
<td>1</td>
<td>2</td>
<td>1.33</td>
</tr>
<tr>
<td>DeRosier, Swick, Ornstein-Davis, Sturtz-McMillen &amp; Matthews (2011)</td>
<td>Weighting High</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td>Score 3</td>
<td>3</td>
<td>1</td>
<td>2.33</td>
</tr>
<tr>
<td></td>
<td>Score 2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Wood et al. (2009)</td>
<td>Weighting Low</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td>Score 1</td>
<td>2</td>
<td>2</td>
<td>1.67</td>
</tr>
</tbody>
</table>
APPENDIX E1


Coding Protocol

Name of Coder: X Date: 10 Feb 2014

Full Study Reference in proper format:


Intervention Name (description of study): Cognitive-behavior therapy-based social skills intervention

Study ID Number: 1

☐ Type of Publication:
  ☑ Book/Monograph
  ☑ Journal Article
  ☑ Book Chapter
  ☐ Other (specify):

1. General Characteristics

A. General Design Characteristics

A1. Random assignment designs (if random assignment design, select one of the following)
  ☑ Completely randomized design
  ☐ Randomized block design (between participants, e.g., matched classrooms)
  ☐ Randomized block design (within participants)
  ☐ Randomized hierarchical design (nested treatments)

A2. Nonrandomized designs (if non-random assignment design, select one of the following)
  ☐ Nonrandomized design
  ☐ Nonrandomized block design (between participants)
  ☐ Nonrandomized block design (within participants)
  ☐ Nonrandomized hierarchical design
  ☐ Optional coding for Quasi-experimental designs
A3. Overall confidence of judgment on how participants were assigned
(select one of the following)

- Very low (little basis)
- Low (guess)
- Moderate (weak inference)
- High (strong inference)
- Very high (explicitly stated)
- N/A
- Unknown/unable to code

B Participants

Total size of sample (start of study): 17

Intervention group sample size: 7

Control group sample size: 8

Attrition: 2 boys from intervention group – one boy dropped out before intervention began; another boy dropped out after participating in 5 of the 15 sessions.

C. Type of Program

- Universal prevention program
- Selective prevention program
- Targeted prevention program
- Intervention/Treatment
- Unknown

D. Stage of Program

- Model/demonstration programs
- Early stage programs
- Established/institutionalized programs
- Unknown

E. Concurrent or Historical Intervention Exposure

- Current exposure
- Prior exposure
- Unknown
Section 2   Key Features for Coding Studies and Rating Level of Evidence/Support

A Measurement (Estimating the quality of the measures used to establish effects)
(Rating Scale: 3= Strong Evidence, 2=Promising Evidence, 1=Weak Evidence, 0=No Evidence)

A1 The use of the outcome measures produce reliable scores for the majority of the primary outcomes (see following table for a detailed breakdown on the outcomes)
- Yes
- No
- Unknown/unable to code

A2 Multi-method (at least two assessment methods used)
- Yes
- No
- N/A
- Unknown/unable to code

A3 Multi-source (at least two sources used self-reports, teachers etc.)
- Yes
- No
- N/A
- Unknown/unable to code

A4 Validity of measures reported (well-known or standardized or norm-referenced are considered good, consider any cultural considerations)
- Yes validated with specific target group
- In part, validated for general population only
- No
- Unknown/unable to code

Overall Rating of Measurement: ☒ 3 ☐ 2 ☐ 1 ☐ 0
### B Comparison Group

#### B1 Type of Comparison group

- Typical intervention
- Attention placebo
- Intervention element placebo
- Alternative intervention
- Pharmacotherapy
- No intervention
- Wait list/delayed intervention
- Minimal contact
- Unable to identify type of comparison
- No comparison group

#### B2 Overall rating of judgment of type of comparison group

- Very low
- Low
- Moderate
- High
- Very high
- Unable to identify comparison group
- Not applicable

#### B3 Counterbalancing of change agent (participants who receive intervention from a single therapist/teacher etc. were counter-balanced across intervention)

- By change agent
- Statistical (analyse includes a test for intervention)
- Other
- Not reported/None

#### B4 Group equivalence established

- Random assignment
- Posthoc matched set
- Statistical matching
- Post hoc test for group equivalence

#### B5 Equivalent mortality

- Low attrition (less than 20 % for post)
- Low attrition (less than 30% for follow-up)
- Intent to intervene analysis carried out?
  
  Findings_____________

### Overall Level of Evidence: 

- **3** = Strong Evidence
- **2** = Promising Evidence
- **1** = Weak Evidence
- **0** = No Evidence
C Primary / Secondary Outcomes

C1 Evidence of appropriate statistical analysis for Primary Outcomes

☐ Appropriate unit of analysis
☐ Familywise/expermenter wise error rate controlled when applicable
☐ Sufficiently large N

C2 Percentage of Primary Outcomes that are significant

Proportion of significant primary outcomes out of the total primary outcome measure for each key construct

☐ At least 75%
☒ 50 – 74%
☐ 25 – 49%
☐ less than 25%

C3 Evidence of appropriate statistical analysis for Secondary Outcomes

☐ Appropriate unit of analysis
☐ Familywise/expermenter wise error rate controlled when applicable
☐ Sufficiently large N
☒ No secondary outcomes reported

C3 Percentage of Secondary Outcomes that are significant

Proportion of significant primary outcomes out of the total secondary outcome measure for each key construct

☐ At least 75%
☐ 50 – 74%
☐ 25 – 49%
☐ less than 25%
☒ Not applicable as no secondary outcomes reported

Overall Rating of Analysis: ☐ 3 ☒ 2 ☐ 1 ☐ 0

3= Strong Evidence 2=Promising Evidence 1=Weak Evidence 0= No Evidence
<table>
<thead>
<tr>
<th>Significant Outcomes</th>
<th>Primary vs Secondary</th>
<th>Who Changed</th>
<th>What Changed</th>
<th>Source</th>
<th>Treatment Information</th>
<th>Outcome Measure Used</th>
<th>Reliability</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome 1</td>
<td></td>
<td>Child</td>
<td>Behaviour</td>
<td>Self Report</td>
<td>Intervention Group vs Control Group</td>
<td>Vineland Adaptive Behavior scales – 2nd Edition (Sparrow, Cicchetti &amp; Balla, 2005)</td>
<td>Standardized Norm-referenced Internal consistency (ages 10 – 12) range from .89 to .92 Test-retest reliability = .93 Partial eta squared = 0.01 Small</td>
<td></td>
</tr>
<tr>
<td>Socialization Skills</td>
<td>Unknown</td>
<td>Teacher</td>
<td>Attitude</td>
<td>Parent Report</td>
<td></td>
<td>Parent / Caregiver Rating Form (Socialization scale) Internal Consistency for boys = .93 Partial eta squared = 0.00 Small</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>Other</td>
<td>Knowledge</td>
<td>Teacher Report</td>
<td></td>
<td>Test-retest reliability = .85 Inter-rater reliability ranging from .75 between father and teacher to .91 between mother father</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>Other</td>
<td>Observation</td>
<td>Observation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Outcome 2

**Social Perception**

*(a) Emotion recognition (Recognising feelings and emotions of self and others)*

| Primary | Secondary | Unknown | Child | Teacher | Parent/Sig.A | Ecology | Other | Unknown | Self Report | Parent Report | Teacher Report | Observation | Test | Other | Unknown | Child and Adolescent Social Perception Measure (CASP; Magill-Evans et al., 1995) Emotion Score measure Internal consistency from .88 to .92 Cronbach α = .75 Test-retest reliability between .83 to .87 Inter-rater reliability from .94 to .99

| Internal consistency from .88 to .92 |
| Partial eta squared = 0.61 |
| Large |

*(b) Use of contextual cues to infer emotions*  

| Primary | Secondary | Unknown | Child | Teacher | Parent/Sig.A | Ecology | Other | Unknown | Self Report | Parent Report | Teacher Report | Observation | Test | Other | Unknown | Nonverbal Cues Score measure Cronbach α = .87 Partial eta squared = 0.45 |
| Partial eta squared = 0.45 |
| Large |

### Outcome 3

**Peer Interaction (Initiating and maintaining conversations)**

| Primary | Secondary | Unknown | Child | Teacher | Parent/Sig.A | Ecology | Other | Unknown | Self Report | Parent Report | Teacher Report | Observation | Test | Other | Unknown | Intervention Group vs Control Group Peer Interaction Measure (PIM; Koning, Magill-Evans, & Volden, 2008) Coding system derived from ADOS (Lord et al., 1999), Magill (1987), and Barry et al. (2003) Cronbach α = .89 Inter-rater reliability = 76% |
| Inter-rater reliability = 76% |
| Large |

| Partial eta squared = 0.29 |
| Large |
|---------------------------|------------------|-----------|---------|----------------|------------------|--------------|---------|----------------|---------|-----------------|------------------|---------------|-----------------|-------------|------------------|--------------|-------------|----------------------|-------|
| Social Knowledge          |                  |           |         |                |                  |              |         |                |         |                 |                  |               |                 |             |                  |              |             | 0.68                  |       |
| (Understanding of social  |                  |           |         |                |                  |              |         |                |         |                 |                  |               |                 |             |                  |              |             | Large                  |       |
| norms and rules;         |                  |           |         |                |                  |              |         |                |         |                 |                  |               |                 |             |                  |              |             | strategies for dealing with difficult situations) |       |
Coding Protocol

Name of Coder: X Date: 10 Feb 2014

Full Study Reference in proper format:

Intervention Name (description of study): Social competence intervention

Study ID Number: 2

☐ Type of Publication:
☐ Book/Monograph
☒ Journal Article
☐ Book Chapter
☐ Other (specify):

1. General Characteristics

A. General Design Characteristics

A1. Random assignment designs (if random assignment design, select one of the following)
☐ Completely randomized design
☒ Randomized block design (between participants, e.g., matched classrooms)
☐ Randomized block design (within participants)
☐ Randomized hierarchical design (nested treatments)

A2. Nonrandomized designs (if non-random assignment design, select one of the following)
☐ Nonrandomized design
☐ Nonrandomized block design (between participants)
☐ Nonrandomized block design (within participants)
☐ Nonrandomized hierarchical design
☒ Optional coding for Quasi-experimental designs
  ➔ No control group – Within-subjects design with pre- and post-tests.
A3. Overall confidence of judgment on how participants were assigned (select one of the following)
- Very low (little basis)
- Low (guess)
- Moderate (weak inference)
- High (strong inference)
- Very high (explicitly stated)
- N/A
- Unknown/unable to code

B Participants

Total size of sample (start of study): 20

Intervention group sample size: 20

No control group

C. Type of Program

- Universal prevention program
- Selective prevention program
- Targeted prevention program
- Intervention/Treatment
- Unknown

D. Stage of Program

- Model/demonstration programs
- Early stage programs
- Established/institutionalized programs
- Unknown

E. Concurrent or Historical Intervention Exposure

- Current exposure
- Prior exposure
- Unknown
Section 2  Key Features for Coding Studies and Rating Level of Evidence/Support

A Measurement (Estimating the quality of the measures used to establish effects)
(Rating Scale: 3= Strong Evidence, 2=Promising Evidence, 1=Weak Evidence, 0=No Evidence)

A1 The use of the outcome measures produce reliable scores for the majority of the primary outcomes (see following table for a detailed breakdown on the outcomes)
- Yes
- No
- Unknown/unable to code for 2 out of 6 outcome measures

A2 Multi-method (at least two assessment methods used)
- Yes
- No
- N/A
- Unknown/unable to code

A3 Multi-source (at least two sources used self-reports, teachers etc.)
- Yes
- No
- N/A
- Unknown/unable to code

A4 Validity of measures reported (well-known or standardized or norm-referenced are considered good, consider any cultural considerations)
- Yes validated with specific target group
- In part, validated for general population only
- No
- Unknown/unable to code

Overall Rating of Measurement: [ ] 3 [ ] 2 [ ] 1 [ ] 0

The overall rating of measurement “2” was awarded despite the study not meeting the requirements of A1 because the test-retest reliability information of one of the measures used in the study, The Reading the Mind in Eyes test, is typically not reported in literature. This is because the measure, like tests explicitly designed to test emotion recognition has psychometric properties that prevent straightforward calculation of Cronbach’s alpha (Fernandez-Abascal, E.G et al., 2013). Although tests like Theory of Mind (TOM) and emotion recognition measures have been most consistently utilised in the social cognition literature, there has been a paucity of measures validated for intervention purposes in this area (Beaumont and Sofronoff, 2008).
B Comparison Group

B1 Type of Comparison group
☐ Typical intervention
☐ Attention placebo
☐ Intervention element placebo
☐ Alternative intervention
☐ Pharmacotherapy
☐ No intervention
☐ Wait list/delayed intervention
☐ Minimal contact
☐ Unable to identify type of comparison
☒ No comparison group

B2 Overall rating of judgment of type of comparison group
☐ Very low
☐ Low
☐ Moderate
☐ High
☐ Very high
☐ Unable to identify comparison group
☒ Not applicable

B3 Counterbalancing of change agent (participants who receive intervention from a single therapist/teacher etc. were counter-balanced across intervention)
☐ By change agent
☐ Statistical (analyse includes a test for intervention)
☐ Other
☒ Not reported/None

B4 Group equivalence established
☐ Random assignment
☐ Posthoc matched set
☐ Statistical matching
☐ Post hoc test for group equivalence
☒ Not reported/None

B5 Equivalent mortality
☒ Low attrition (less than 20 % for post)
☐ Low attrition (less than 30% for follow-up)
☐ Intent to intervene analysis carried out?
Findings _______________

Overall Level of Evidence: ☒ 3 ☐ 2 ☐ 1 ☒ 0

3 = Strong Evidence 2 = Promising Evidence 1 = Weak Evidence 0 = No Evidence
C Primary / Secondary Outcomes

C1 Evidence of appropriate statistical analysis for Primary Outcomes

☑ Appropriate unit of analysis
☐ Familywise/expermenter wise error rate controlled when applicable
☐ Sufficiently large N

C2 Percentage of Primary Outcomes that are significant
Proportion of significant primary outcomes out of the total primary outcome measure for each key construct

☐ At least 75%
☑ 50 – 74%
☐ 25 – 49%
☐ less than 25%

C3 Evidence of appropriate statistical analysis for Secondary Outcomes

☐ Appropriate unit of analysis
☐ Familywise/expermenter wise error rate controlled when applicable
☐ Sufficiently large N
☑ No secondary outcomes reported

C3 Percentage of Secondary Outcomes that are significant
Proportion of significant primary outcomes out of the total secondary outcome measure for each key construct

☐ At least 75%
☐ 50 – 74%
☐ 25 – 49%
☐ less than 25%
☑ Not applicable as no secondary outcomes reported

Overall Rating of Analysis: ☐ 3 ☑ 2 ☐ 1 ☐ 0

3= Strong Evidence 2=Promising Evidence 1=Weak Evidence 0= No Evidence
<table>
<thead>
<tr>
<th>Significant Outcomes</th>
<th>Primary vs Secondary</th>
<th>Who Changed</th>
<th>What Changed</th>
<th>Source</th>
<th>Treatment Information</th>
<th>Outcome Measure Used</th>
<th>Reliability</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome 1</td>
<td>Overall Social Abilities</td>
<td>Primary</td>
<td>Secondary</td>
<td>Unknown</td>
<td>Intervention Group</td>
<td>Social Responsiveness Skill (SRS; Constantino &amp; Gruber, 2005)</td>
<td>Standardized Norm-referenced Internal consistency (ages 10 – 12) range from.89 to.92 Test-retest reliability = .93</td>
<td>Social Abilities Cohen’s d = 0.75 (parent measure) Medium Cohen’s d = 0.39 (teacher measure) Small</td>
</tr>
<tr>
<td>(a) Social Awareness</td>
<td></td>
<td>Child</td>
<td>Teacher</td>
<td>Parent/ Sig.A</td>
<td></td>
<td></td>
<td></td>
<td>Small</td>
</tr>
<tr>
<td>(b) Social Cognition</td>
<td></td>
<td>Attitude</td>
<td>Knowledge</td>
<td>Teacher Report</td>
<td></td>
<td></td>
<td></td>
<td>Small</td>
</tr>
<tr>
<td>(c) Social Communication</td>
<td></td>
<td>Other</td>
<td>Unknown</td>
<td>Observation</td>
<td></td>
<td></td>
<td></td>
<td>Small</td>
</tr>
<tr>
<td>(d) Social Motivation</td>
<td></td>
<td>Other</td>
<td>Unknown</td>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td>Small</td>
</tr>
</tbody>
</table>

Cohen’s d = 0.33 (parent measure) Small
Cohen’s d = 0.31 (teacher measure) Small

Cohen’s d = 0.61 (parent measure) Medium
Cohen’s d = 0.27 (teacher measure) Small

Cohen’s d = 0.74 (parent measure) Medium
Cohen’s d = 0.40 (teacher measure) Small

Cohen’s d = 0.60 (parent measure) Medium
Cohen’s d = 0.41 (teacher measure) Small
### Outcome 2
**Theory of Mind (ToM)**
- **Perspective taking**

<table>
<thead>
<tr>
<th>Primary</th>
<th>Secondary</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child</td>
<td>Teacher</td>
<td>Parent/Sig.A</td>
</tr>
<tr>
<td>Ecoty</td>
<td>Other</td>
<td>Unknown</td>
</tr>
<tr>
<td>Behaviour</td>
<td>Attitude</td>
<td>Knowledge</td>
</tr>
<tr>
<td>Other</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>Observation</td>
<td>Test</td>
<td>Other</td>
</tr>
<tr>
<td>Other</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

**First Order ToM**
- Sally-Anne False Belief Task (Baron-Cohen et al., 1985)
- Smurp False Belief Task (Perner et al., 1989)
- Faux Pas Stories (Baron-Cohen et al., 1999)

- Insufficient data to calculate Cohen's $d$ = 0.40
- Small

### Outcome 3
**Emotion Recognition**

<table>
<thead>
<tr>
<th>Primary</th>
<th>Secondary</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child</td>
<td>Teacher</td>
<td>Parent/Sig.A</td>
</tr>
<tr>
<td>Ecology</td>
<td>Other</td>
<td>Unknown</td>
</tr>
<tr>
<td>Behaviour</td>
<td>Attitude</td>
<td>Knowledge</td>
</tr>
<tr>
<td>Other</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>Observation</td>
<td>Test</td>
<td>Other</td>
</tr>
<tr>
<td>Other</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

**Intervention Group**
- Diagnostic Analysis of Non-Verbal Accuracy – 2, Child Facial Expressions (DANVA-2-CF; Nowicki and Carton, 1993)
- Reading the Mind in Eyes Test (Baron-Cohen et al. 2001)

- Internal consistency from .69 to .81
- Test-retest reliability = .74
- Not reported by authors
- Cohen's $d$ = 0.13
- Small
<table>
<thead>
<tr>
<th>Outcome 4</th>
<th>Executive Functioning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- Metacognition</td>
</tr>
<tr>
<td></td>
<td>- Behaviour Regulation (Emotional Control)</td>
</tr>
<tr>
<td></td>
<td>- Problem Solving</td>
</tr>
<tr>
<td>Primary</td>
<td>Secondary</td>
</tr>
<tr>
<td>Function</td>
<td>Unknown</td>
</tr>
<tr>
<td>- Child</td>
<td>Teacher</td>
</tr>
<tr>
<td>- Parent/Sig.A</td>
<td>Ecology</td>
</tr>
<tr>
<td>- Other</td>
<td>Unknown</td>
</tr>
<tr>
<td>- Self Report</td>
<td>Parent Report</td>
</tr>
<tr>
<td>- Teacher Report</td>
<td>Knowledge</td>
</tr>
<tr>
<td>- Other</td>
<td>Unknown</td>
</tr>
<tr>
<td>- Test</td>
<td>Unknown</td>
</tr>
<tr>
<td>Intervention Group</td>
<td>Behaviour Rating Inventory of Executive Function (BRIEF: Gioia et al., 2000)</td>
</tr>
<tr>
<td>- Metacognition index</td>
<td>Internal consistency from .80 to .98</td>
</tr>
<tr>
<td></td>
<td>Test-retest reliability = .82 for parents</td>
</tr>
<tr>
<td>- Behaviour Regulation index</td>
<td>Cohen’s $d$ = 0.42</td>
</tr>
<tr>
<td></td>
<td>Small</td>
</tr>
<tr>
<td>Test of Problem Solving – 3 (TOPS-3: Elementary; Bowers et al., 2005)</td>
<td>Test-retest reliability = .84</td>
</tr>
<tr>
<td></td>
<td>Cohen’s $d$ = 0.15</td>
</tr>
<tr>
<td></td>
<td>Small</td>
</tr>
</tbody>
</table>
Coding Protocol

Name of Coder: X

Date: 10 Feb 2014

Full Study Reference in proper format:


Intervention Name (description of study): Social skills group intervention for children with high functioning autism spectrum disorders

Study ID Number: 4

- Type of Publication:
  - Book/Monograph
  - Journal Article
  - Book Chapter
  - Other (specify):

1. General Characteristics

A. General Design Characteristics

A1. Random assignment designs (if random assignment design, select one of the following)

- Completely randomized design
- Randomized block design (between participants, e.g., matched classrooms)
- Randomized block design (within participants)
- Randomized hierarchical design (nested treatments)

A2. Nonrandomized designs (if non-random assignment design, select one of the following)

- Nonrandomized design
- Nonrandomized block design (between participants)
- Nonrandomized block design (within participants)
- Nonrandomized hierarchical design
- Optional coding for Quasi-experimental designs
A3. Overall confidence of judgment on how participants were assigned (select one of the following)

- Very low (little basis)
- Low (guess)
- Moderate (weak inference)
- High (strong inference)
- Very high (explicitly stated)
- N/A
- Unknown/unable to code

B Participants

Total size of sample (start of study): 55

Intervention group sample size: 27

Control group sample size: 28

C. Type of Program

- Universal prevention program
- Selective prevention program
- Targeted prevention program
- Intervention/Treatment
- Unknown

D. Stage of Program

- Model/demonstration programs
- Early stage programs
- Established/institutionalized programs
- Unknown

E. Concurrent or Historical Intervention Exposure

- Current exposure
- Prior exposure
- Unknown
Section 2  Key Features for Coding Studies and Rating Level of Evidence/Support

A Measurement (Estimating the quality of the measures used to establish effects)
(Rating Scale: 3= Strong Evidence, 2=Promising Evidence, 1=Weak Evidence, 0=No Evidence)

A1 The use of the outcome measures produce reliable scores for the majority of the primary outcomes (see following table for a detailed breakdown on the outcomes)

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
<th>Unknown/unable to code</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>☒</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

A2 Multi-method (at least two assessment methods used)

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
<th>N/A</th>
<th>Unknown/unable to code</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>☒</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

A3 Multi-source (at least two sources used self-reports, teachers etc.)

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
<th>N/A</th>
<th>Unknown/unable to code</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>☒</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

A4 Validity of measures reported (well-known or standardized or norm-referenced are considered good, consider any cultural considerations)

<table>
<thead>
<tr>
<th></th>
<th>Yes validated with specific target group</th>
<th>In part, validated for general population only</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>☒</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

Overall Rating of Measurement: ☒ 3 ☐ 2 ☐ 1 ☐ 0
B Comparison Group

B1 Type of Comparison group
- Typical intervention
- Attention placebo
- Intervention element placebo
- Alternative intervention
- Pharmacotherapy
- No intervention
- Wait list/delayed intervention
- Minimal contact
- Unable to identify type of comparison
- No comparison group

B2 Overall rating of judgment of type of comparison group
- Very low
- Low
- Moderate
- High
- Very high
- Unable to identify comparison group
- Not applicable

B3 Counterbalancing of change agent (participants who receive intervention from a single therapist/teacher etc. were counter-balanced across intervention)
- By change agent
- Statistical (analyse includes a test for intervention)
- Other
- Not reported/None

B4 Group equivalence established
- Random assignment
- Posthoc matched set
- Statistical matching
- Post hoc test for group equivalence
- Not reported/None

B5 Equivalent mortality
- Low attrition (less than 20 % for post)
- Low attrition (less than 30% for follow-up)

Findings__________________

Overall Level of Evidence: 3 2 1 0

3= Strong Evidence 2=Promising Evidence 1=Weak Evidence 0= No Evidence
C Primary / Secondary Outcomes

C1 Evidence of appropriate statistical analysis for Primary Outcomes
  □ Appropriate unit of analysis
  □ Familywise/expermenter wise error rate controlled when applicable
  □ Sufficiently large N

C2 Percentage of Primary Outcomes that are significant
  Proportion of significant primary outcomes out of the total primary outcome measure for each key construct
  □ At least 75%
  □ 50 – 74%
  □ 25 – 49%
  □ less than 25%

C3 Evidence of appropriate statistical analysis for Secondary Outcomes
  □ Appropriate unit of analysis
  □ Familywise/expermenter wise error rate controlled when applicable
  □ Sufficiently large N
  □ No secondary outcomes reported

C3 Percentage of Secondary Outcomes that are significant
  Proportion of significant primary outcomes out of the total secondary outcome measure for each key construct
  □ At least 75%
  □ 50 – 74%
  □ 25 – 49%
  □ less than 25%
  □ Not applicable as no secondary outcomes reported

Overall Rating of Analysis: □ 3 □ 2 □ 1 □ 0
3= Strong Evidence    2=Promising Evidence    1=Weak Evidence    0= No Evidence
## Significant Outcomes

<table>
<thead>
<tr>
<th>Outcome 1</th>
<th>Primary vs Secondary</th>
<th>Who Changed</th>
<th>What Changed</th>
<th>Source</th>
<th>Treatment Information</th>
<th>Outcome Measure Used</th>
<th>Reliability</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Social Abilities</td>
<td>Primary</td>
<td>Child</td>
<td>Behaviour</td>
<td>Self Report</td>
<td>Intervention</td>
<td>Achieved</td>
<td>Cronbach α = .79</td>
<td>Cohen’s $d$ = 0.75</td>
</tr>
<tr>
<td>Secondary</td>
<td>Teacher</td>
<td>Attitude</td>
<td>Parent Report</td>
<td></td>
<td>Group vs Control Group</td>
<td>Learning Questionnaire (ALQ; DeRosier &amp; Gilliom, 2007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>Parent/Sig.A</td>
<td>Knowledge</td>
<td>Teacher Report</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ecology</td>
<td>Other</td>
<td>Observation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>Unknown</td>
<td>Other</td>
<td>Test</td>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(e) Social Awareness
(f) Social Cognition
(g) Social Communication
(h) Social Motivation

Social Responsiveness Skill (SRS; Constantino & Gruber, 2005)
Cronbach α = .80

(e) Social awareness
Cohen’s $d$ = 0.69 **Medium**

(f) Social cognition
Cohen’s $d$ was not reported as there were no differences between means found

(g) Social communication
Cohen’s $d$ = 0.94 **Large**

(h) Social motivation
Cohen’s $d$ = 0.87 **Medium**
<table>
<thead>
<tr>
<th>Outcome 2</th>
<th>Social Self-Efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>Secondary</td>
</tr>
<tr>
<td>Unknown</td>
<td></td>
</tr>
</tbody>
</table>

- Child
- Teacher
- Parent/Sig.A
- Ecology
- Other
- Unknown

- Behaviour
- Attitude
- Knowledge
- Other
- Unknown

- Self Report
- Parent Report
- Teacher Report
- Observation
- Test
- Other
- Unknown

- Intervention Group vs Control Group
- Social Self-Efficacy Scale (Ollendick & Schmidt, 1987)

- Internal consistency reliability = .87
- Test-retest reliability = .75

Cronbach α = .81 for current sample

Cohen’s $d$ was not reported as there were no differences between means found.
Coding Protocol

Name of Coder: X  
Date: 10 Feb 2014

Full Study Reference in proper format:


Intervention Name (description of study): Manualized social treatment for high-functioning autism spectrum disorders

Study ID Number: 3

Type of Publication:

- [ ] Book/Monograph
- [x] Journal Article
- [ ] Book Chapter
- [ ] Other (specify):

1. General Characteristics

A. General Design Characteristics

A1. Random assignment designs (if random assignment design, select one of the following)

- [x] Completely randomized design
- [ ] Randomized block design (between participants, e.g., matched classrooms)
- [ ] Randomized block design (within participants)
- [ ] Randomized hierarchical design (nested treatments)

A2. Nonrandomized designs (if non-random assignment design, select one of the following)

- [ ] Nonrandomized design
- [ ] Nonrandomized block design (between participants)
- [ ] Nonrandomized block design (within participants)
- [ ] Nonrandomized hierarchical design
- [ ] Optional coding for Quasi-experimental designs
A3. Overall confidence of judgment on how participants were assigned
(select one of the following)
☐ Very low (little basis)
☐ Low (guess)
☐ Moderate (weak inference)
☐ High (strong inference)
☒ Very high (explicitly stated)
☐ N/A
☐ Unknown/unable to code

B Participants

Total size of sample (start of study): 36

Intervention group sample size: 18

Control group sample size: 18

C. Type of Program

☐ Universal prevention program
☐ Selective prevention program
☐ Targeted prevention program
☒ Intervention/Treatment
☐ Unknown

D. Stage of Program

☐ Model/demonstration programs
☐ Early stage programs
☒ Established/institutionalized programs
☐ Unknown

E. Concurrent or Historical Intervention Exposure

☐ Current exposure
☐ Prior exposure
☒ Unknown
Section 2  Key Features for Coding Studies and Rating Level of Evidence/Support

A Measurement (Estimating the quality of the measures used to establish effects)
(Rating Scale: 3= Strong Evidence, 2=Promising Evidence, 1=Weak Evidence, 0=No Evidence)

A1 The use of the outcome measures produce reliable scores for the majority of the primary outcomes (see following table for a detailed breakdown on the outcomes)
- Yes
- No
- Unknown/unable to code

A2 Multi-method (at least two assessment methods used)
- Yes
- No
- N/A
- Unknown/unable to code

A3 Multi-source (at least two sources used self-reports, teachers etc.)
- Yes
- No
- N/A
- Unknown/unable to code

A4 Validity of measures reported (well-known or standardized or norm-referenced are considered good, consider any cultural considerations)
- Yes validated with specific target group
- In part, validated for general population only
- No
- Unknown/unable to code

Overall Rating of Measurement: 3 2 1 0
B Comparison Group

B1 Type of Comparison group

- Typical intervention
- Attention placebo
- Intervention element placebo
- Alternative intervention
- Pharmacotherapy
- No intervention
- Wait list/delayed intervention
- Minimal contact
- Unable to identify type of comparison
- No comparison group

B2 Overall rating of judgment of type of comparison group

- Very low
- Low
- Moderate
- High
- Very high
- Unable to identify comparison group
- Not applicable

B3 Counterbalancing of change agent (participants who receive intervention from a single therapist/teacher etc. were counter-balanced across intervention)

- By change agent
- Statistical (analyse includes a test for intervention)
- Other
- Not reported/None

B4 Group equivalence established

- Random assignment
- Posthoc matched set
- Statistical matching
- Post hoc test for group equivalence
- Not reported/None

B5 Equivalent mortality

- Low attrition (less than 20 % for post)
- Low attrition (less than 30% for follow-up)
- Intent to intervene analysis carried out?
  Findings________________

Overall Level of Evidence: □ 3  □  2  □ 1 □ 0

3= Strong Evidence  2=Promising Evidence  1=Weak Evidence  0= No Evidence
C Primary / Secondary Outcomes

C1 Evidence of appropriate statistical analysis for Primary Outcomes

- Appropriate unit of analysis
- Familywise/expermenter wise error rate controlled when applicable
- Sufficiently large N

C2 Percentage of Primary Outcomes that are significant
Proportion of significant primary outcomes out of the total primary outcome measure for each key construct

- At least 75%
- 50 – 74%
- 25 – 49%
- less than 25%

C3 Evidence of appropriate statistical analysis for Secondary Outcomes

- Appropriate unit of analysis
- Familywise/expermenter wise error rate controlled when applicable
- Sufficiently large N
- No secondary outcomes reported

C3 Percentage of Secondary Outcomes that are significant
Proportion of significant primary outcomes out of the total secondary outcome measure for each key construct

- At least 75%
- 50 – 74%
- 25 – 49%
- less than 25%
- Not applicable as no secondary outcomes reported

Overall Rating of Analysis: ☐ 3  ☑ 2  ☐ 1  ☐ 0

3= Strong Evidence  2=Promising Evidence  1=Weak Evidence  0= No Evidence
## Significant Outcomes

**Outcome 1**

**Overall Social Abilities**

<table>
<thead>
<tr>
<th>Primary vs Secondary</th>
<th>Who Changed</th>
<th>What Changed</th>
<th>Source</th>
<th>Treatment Information</th>
<th>Outcome Measure Used</th>
<th>Reliability</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>Child</td>
<td>Behaviour</td>
<td>Self Report</td>
<td>Intervention Group vs Control Group</td>
<td>Social Responsiveness Skill (SRS; Constantino &amp; Gruber, 2005)</td>
<td>Standardized Norm-referenced Internal consistency reliability range from .93 to .97, discriminates between spectrum and non-spectrum behavioural disorders</td>
<td>Cohen’s $d = 0.625$ (parent measure) Medium</td>
</tr>
<tr>
<td>Secondary</td>
<td>Teacher</td>
<td>Attitude</td>
<td>Parent Report</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>Parent/Sig.A</td>
<td>Knowledge</td>
<td>Teacher Report</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
<td>Other</td>
<td>Observation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>Unknown</td>
<td>Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
<td>Unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cohen’s $d = 0.711$ (teacher measure) **Medium**

**Adapted Skillstreaming Checklist (ASC; Lopata et al., 2008)**

- Designed as a direct measure of skills taught in the social program

Cohen’s $d = 0.584$ (parent measure) **Medium**

Cohen’s $d = 1.421$ (teacher measure) **Large**
### Outcome 2

**Positive Social Response**

(a) Reduced tendency to escape from or avoid social contact

<table>
<thead>
<tr>
<th>Primary</th>
<th>Secondary</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child</td>
<td>Teacher</td>
<td>Parent/ Sig. A</td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
<td>Other</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Behaviour</th>
<th>Attitude</th>
<th>Knowledge</th>
<th>Other</th>
<th>Unknown</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Self Report</th>
<th>Parent Report</th>
<th>Teacher Report</th>
<th>Observation</th>
<th>Test</th>
<th>Other</th>
<th>Unknown</th>
</tr>
</thead>
</table>


- Internal consistency reliability range from .77 to .88 (PRS) and .82 to .93 (TRS)
- Withdrawal
  - Cohen's $d$ = 1.055 (parent measure)
  - Large

- Social Skills
  - Cohen's $d$ = 0.693 (teacher measure)
  - Medium

(b) Positive social functioning

2 scales were used:
Withdrawal and Social Skills

- Withdrawal
  - Cohen's $d$ = 1.272 (teacher measure)
  - Large

### Outcome 3

**Social Knowledge**

<table>
<thead>
<tr>
<th>Primary</th>
<th>Secondary</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child</td>
<td>Teacher</td>
<td>Parent/ Sig. A</td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
<td>Other</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Behaviour</th>
<th>Attitude</th>
<th>Knowledge</th>
<th>Other</th>
<th>Unknown</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Self Report</th>
<th>Parent Report</th>
<th>Teacher Report</th>
<th>Observation</th>
<th>Test</th>
<th>Other</th>
<th>Unknown</th>
</tr>
</thead>
</table>

Intervention Group vs Control Group

Skillstreaming Knowledge Assessment (SKA; Lopata et al., 2010)

- Internal consistency reliability = .94
- Inter-rater reliability coefficient = .75

Cohen's $d$ = 1.272 (teacher measure)

Large
<table>
<thead>
<tr>
<th>Outcome 4</th>
<th>Understanding of Idiomatic Language</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary</strong></td>
<td>Child</td>
</tr>
<tr>
<td><strong>Secondary</strong></td>
<td>Teacher</td>
</tr>
<tr>
<td><strong>Unknown</strong></td>
<td>Parent/Sig.A</td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
</tr>
<tr>
<td></td>
<td>Other</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
</tr>
<tr>
<td><strong>Intervention</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Group vs Control Group</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Intervention</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Group</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Cohen’s d</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Self Report</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Parent Report</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Teacher Report</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Observation</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Test</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Unknown</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome 5</th>
<th>Emotion Recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary</strong></td>
<td>Child</td>
</tr>
<tr>
<td><strong>Secondary</strong></td>
<td>Teacher</td>
</tr>
<tr>
<td><strong>Unknown</strong></td>
<td>Parent/Sig.A</td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
</tr>
<tr>
<td></td>
<td>Other</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
</tr>
<tr>
<td><strong>Intervention</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Group vs Control Group</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Intervention</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Group</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Cohen’s d</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Self Report</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Parent Report</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Teacher Report</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Observation</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Test</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Unknown</strong></td>
<td></td>
</tr>
</tbody>
</table>
Cognitive behavioural therapy for school-aged children with high functioning autism

Study ID Number: 5

1. General Characteristics

A. General Design Characteristics

A1. Random assignment designs (if random assignment design, select one of the following)
- Completely randomized design
- Randomized block design (between participants, e.g., matched classrooms)
- Randomized hierarchical design (nested treatments)

A2. Nonrandomized designs (if non-random assignment design, select one of the following)
- Nonrandomized design
- Nonrandomized block design (between participants)
- Nonrandomized block design (within participants)
- Nonrandomized hierarchical design
- Optional coding for Quasi-experimental designs
A3. Overall confidence of judgment on how participants were assigned (select one of the following)
- Very low (little basis)
- Low (guess)
- Moderate (weak inference)
- High (strong inference)
- Very high (explicitly stated)
- N/A
- Unknown/unable to code

B Participants
Total size of sample (start of study): 19
Intervention group sample size: 9
Waitlist group sample size: 10

C. Type of Program
- Universal prevention program
- Selective prevention program
- Targeted prevention program
- Intervention/Treatment
- Unknown

D. Stage of Program
- Model/demonstration programs
- Early stage programs
- Established/institutionalized programs
- Unknown

E. Concurrent or Historical Intervention Exposure
- Current exposure
- Prior exposure → 11 participants were previously from the Wood et al. (2009) trial
- Unknown
Section 2  Key Features for Coding Studies and Rating Level of Evidence/Support

A Measurement (Estimating the quality of the measures used to establish effects)
(Rating Scale: 3= Strong Evidence, 2=Promising Evidence, 1=Weak Evidence, 0=No Evidence)

A1 The use of the outcome measures produce reliable scores for the majority of the primary outcomes (see following table for a detailed breakdown on the outcomes)
- Yes
- No
- Unknown/unable to code

A2 Multi-method (at least two assessment methods used)
- Yes
- No
- N/A
- Unknown/unable to code

A3 Multi-source (at least two sources used self-reports, teachers etc.)
- Yes
- No
- N/A
- Unknown/unable to code

A4 Validity of measures reported (well-known or standardized or norm-referenced are considered good, consider any cultural considerations)
- Yes validated with specific target group
- In part, validated for general population only
- No
- Unknown/unable to code

Overall Rating of Measurement: ☒ 3 ☐ 2 ☐ 1 ☒ 0
B Comparison Group

B1 Type of Comparison group
- Typical intervention
- Attention placebo
- Intervention element placebo
- Alternative intervention
- Pharmacotherapy
- No intervention
- Wait list/delayed intervention
- Minimal contact
- Unable to identify type of comparison
- No comparison group

B2 Overall rating of judgment of type of comparison group
- Very low
- Low
- Moderate
- High
- Very high
- Unable to identify comparison group
- Not applicable

B3 Counterbalancing of change agent (participants who receive intervention from a single therapist/teacher etc. were counter-balanced across intervention)
- By change agent
- Statistical (analyse includes a test for intervention)
- Other → Therapists were randomly assigned
- Not reported/None

B4 Group equivalence established
- Random assignment
- Posthoc matched set
- Statistical matching
- Post hoc test for group equivalence
- Not reported/None

B5 Equivalent mortality
- Low attrition (less than 20% for post)
- Low attrition (less than 30% for follow-up)
- Intent to intervene analysis carried out?
  Findings ______________

Overall Level of Evidence: □ 3 □ 2 □ 1 □ 0
3= Strong Evidence 2=Promising Evidence 1=Weak Evidence 0= No Evidence
C Primary / Secondary Outcomes

C1 Evidence of appropriate statistical analysis for Primary Outcomes

- ✔ Appropriate unit of analysis
- ✔ Familywise/expermenter wise error rate controlled when applicable
- ✔ Sufficiently large N

C2 Percentage of Primary Outcomes that are significant
Proportion of significant primary outcomes out of the total primary outcome measure for each key construct

- ✔ At least 75%
- ☐ 50 – 74%
- ☐ 25 – 49%
- ☐ less than 25%

C3 Evidence of appropriate statistical analysis for Secondary Outcomes

- ☐ Appropriate unit of analysis
- ☐ Familywise/expermenter wise error rate controlled when applicable
- ✔ Sufficiently large N
- ✔ No secondary outcomes reported

C3 Percentage of Secondary Outcomes that are significant
Proportion of significant primary outcomes out of the total secondary outcome measure for each key construct

- ☐ At least 75%
- ☐ 50 – 74%
- ☐ 25 – 49%
- ☐ less than 25%
- ✔ Not applicable as no secondary outcomes reported

Overall Rating of Analysis: 3 2 1 0

3= Strong Evidence 2=Promising Evidence 1=Weak Evidence 0= No Evidence
<table>
<thead>
<tr>
<th>Significant Outcomes</th>
<th>Primary vs Secondary</th>
<th>Who Changed</th>
<th>What Changed</th>
<th>Source</th>
<th>Treatment Information</th>
<th>Outcome Measure Used</th>
<th>Reliability</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall Social Abilities</td>
<td></td>
<td>Child</td>
<td>Behaviour</td>
<td>Self Report</td>
<td>Intervention Group vs Control Group</td>
<td>Social Responsiveness Skill (SRS; Constantino &amp; Gruber, 2005)</td>
<td>Standardized Norm-referenced Cronbach α = .80</td>
<td>Cohen’s $d$ = 0.77</td>
</tr>
<tr>
<td>(i) Social Awareness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(j) Social Cognition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(k) Social Communication</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(l) Social Motivation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(i) Social awareness
Cohen’s $d$ was not reported. Insufficient information provided to calculate effect size.

(j) Social cognition
Cohen’s $d$ was not reported. Insufficient information provided to calculate effect size.

(k) Social communication
Cohen’s $d$ was not reported. Insufficient information provided to calculate effect size.

(l) Social motivation
Cohen’s $d$ was not reported. Insufficient information provided to calculate effect size.