Secret Agent Society – Whole of Classroom (SAS-WOC)Trial

FINAL REPORT

Assoc. Professor Kate Sofronoff
Dr Sharon Hinton
Ruby Clarke
Assistant Professor Renae Beaumont
Dr Jamin Day

March 2019

Australian Government
Department of Industry, Innovation and Science
Business Cooperative Research Centres Program

Secret Agent Society – Whole of Classroom (SAS-WOC) Trial

SAS-WOC Trial

Associate Professor Kate Sofronoff

Associate Professor in Clinical Psychology

School of Psychology, The University of Queensland

Assistant Professor Renae Beaumont

Assistant Professor of Psychology in Clinical Psychiatry

Weill Cornell Medicine, New York Presbyterian Hospital

Dr Sharon Hinton

Research Manager, The University of Queensland

Dr Jamin Day

Postdoctoral Research Fellow

Family Action Centre, The University of Newcastle

Ruby Clarke

University of Queensland

ISBN: 978-1-922365-03-3

Citation: Sofronoff, K., Beaumont., R., Hinton, S., Day, J., & Clarke, R. (2019) Secret Agent Society – Whole of Classroom (SAS-WOC) Trial. Full Report. Brisbane: Cooperative Research Centre for Living with Autism.

Copies of this report can be downloaded from the Autism CRC website **autismcrc.com.au**.

The Cooperative Research Centre for Living with Autism (Autism CRC)

The Cooperative Research Centre for Living with Autism (Autism CRC) is the world's first national, cooperative research effort focused on autism. Taking a whole of life approach to autism focusing on diagnosis, education and adult life, Autism CRC researchers are working with end-users to provide evidence-based outcomes which can be translated into practical solutions for governments, service providers, education and health professionals, families and people on the autism spectrum.

autismcrc.com.au

Acknowledgements

The authors acknowledge the financial support of the Cooperative Research Centre for Living with Autism (Autism CRC), established and supported under the Australian Government's Cooperative Research Centre Program. Staff and non-staff in kind support was provided by Autism CRC participants:

- Rochelle Johnson

The authors would like to thank the teachers, support staff and students who participated in the SAS-WOC program from the following schools:

- Bethany Lutheran College
- Congupna Primary School
- Darling Heights State School
- Diggers Rest Primary School
- Eleebana Public School
- Holy Trinity Catholic Primary
- Livingstone Christian College
- Lumen Christi College
- Murchison Primary School
- Our Lady of the Southern Cross Primary School
- Pacific Paradise State School
- St Augustine's Parish Primary School
- St Mary's Catholic Primary School
- Shepparton East Primary School
- Urangan Point State School
- Varsity College

The authors also wish to acknowledge the additional support and assistance provided by Eleebana Public School, in particular Ms. Amanda Veale, Mr. Ethan Kibble and the students of class 5K (2017). Thank you for allowing us into your classroom to film your personal experiences as participants in SAS-WOC. Your enthusiasm in providing your individual experiences, insights and feedback was greatly appreciated and remains a highlight of the project for the team!

Lastly, the authors would also like to acknowledge the assistance of the following individuals who assisted with questionnaire and data management during the trial:

- Ho Yin Lam
- Sheung Yan Ng
- Sophia Neill
- Kathryn Farr
- Charlotte Qi Song
- Zoe Dale
- Monica McSweeney
- Marilyn Zi Qing Liew
- Merryn Taylor
- Tahlia Rolph
- Jenny Wright
- Ruby Clarke
- Hananeh Abutaleb Kermani

Table of contents

1. Background

1.1 A whole of classroom approach

Autism Spectrum Disorder (ASD) is a life-long developmental disorder, with diagnosis commonly occurring in childhood (American Psychiatric Association, 2013; Simonoff et al., 2008). Autism is diagnosed using a deficit model and is characterized by challenges in social, emotional and behavioral areas of function. Impairment in social functioning is considered a hallmark of autism, embodied as challenges in social communication, social interaction, and maintaining and understanding interpersonal relationships (American Psychiatric Association, 2013; White et al., 2013). People diagnosed with autism may also exhibit inflexibility in terms of routine and patterns of behavior that, when compromised or changed, lead to significant emotional distress (American Psychiatric Association, 2013). This emotional distress is in part due to poor emotion recognition and regulation skills, where the individual has difficulty detecting and understanding emotions within themselves and within others (Loveland, 2005; White et al., 2009). Notably, impairment in social functioning becomes more pronounced during later childhood and adolescence as this developmental period involves an increase in complex social interactions that require more nuanced social and emotional function (Picci & Scherf, 2015; White et al., 2007).

In a mainstream classroom, school success generally requires that students be able to interact positively not only with their peers but also with teachers and other administration and support staff. While it is acknowledged that children on the autism spectrum may have widely varying classroom presentation and experience significant variability in difficulties experienced, it remains that in the very situation that requires the greatest need to understand subtle social communications and protocols their social deficits will be, paradoxically, at their worst (Hinton, Sofronoff & Sheffield, 2008). The rationale for conducting a social-emotional skills program in mainstream classrooms is that this is the ideal context in which to learn and to practice these skills for children on the autism spectrum or for those with other social-emotional difficulties. An expected additional outcome is that a positive portrayal of individual differences will lead not only to skill developmental but also to greater understanding and tolerance of diversity in all of the children involved in the program.

1.1.1 Interventions available for autism

Social skill development is the primary focus of the majority of interventions currently available for autism (White et al., 2007; Gates et al., 2017). This may in part be due to the diagnostic criteria and conceptualisation of the disorder, where impairment in social function is a core component of an autism diagnosis and people on the spectrum are considered to be at increased risk of isolation due to social skill challenges (Mendelson, Gates & Lerner, 2016). Based on the hypotheses that social interaction is underpinned by emotional function (e.g. emotion recognition and emotion regulation; see Attwood, 2004; Beaumont, Rotolone & Sofronoff, 2015; White et al., 2010; White et al., 2013), there has been an increase in interventions that target social and emotional function concurrently. Termed 'social-emotional' interventions, these methods are founded on the basis that both social and emotional domains need to be addressed for effective outcomes (Beaumont & Sofronoff, 2013; Sofronoff, Beaumont & Weiss, 2015). Such an approach is drawn from a Cognitive Behavior Therapy (CBT) perspective, which holds that thoughts, feelings and behaviors are connected (Attwood, 2004; Koning et al., 2013; White et al., 2013). The CBT foundation enables comorbid presentations such as anxiety to be addressed (Attwood, 2004; Beaumont & Sofronoff, 2008; White et al., 2013; Sofronoff et al., 2015). Given the tendency for impairment in social functioning to become more pronounced during later childhood and adolescence (Picci & Scherf, 2015; White et al., 2007), a strong focus is placed on targeting youth (people aged less than 18 years) with social skill intervention (Koning et al., 2013; White et al., 2007; Gates et al., 2017).

To date, interventions for autism have been found to produce mixed results, with study design reported to be of mixed quality (see Gates, Kang & Lerner, 2017; Lounds et al., 2012; Reichow & Volkmar, 2010). This is likely to be due to initial gains made during intervention failing to endure over time and/or to generalise to other contexts outside intervention delivery (Gates et al., 2017; Reichow & Volkmar, 2010; White et al., 2007). A recent finding indicating that the presence of a comorbid anxiety disorder can reduce the effects of social skill interventions delivered to children on the spectrum (Pellecchia, et al., 2015) suggests that interventions should also consider, if not attempt to ameliorate, the effects of comorbid disorders in addition to improving social skills.

1.1.2 The Secret Agent Society: Social-Emotional Skills Program

The Secret Agent Society (SAS) program was developed as a small group program for children (aged 8 to 12 years) with a diagnosis of autism spectrum disorder or Asperger syndrome who

were functioning at a level that ensured they were cognitively able to understand the content of the program. The SAS program uses a secret agent themed approach and incorporates a computer game through which children can learn to identify social and emotional clues and cues that can inform their behavior going forward. During small group sessions children practice and role play the skills and strategies that are taught, while parents are also taught the strategies and asked to encourage their children to practice in other settings and to provide rewards for efforts made. Tip sheets are provided for teachers to further encourage generalization to other settings. The SAS program meets the guidelines for programs created specifically for children on the autism spectrum (Moree & Davis, 2010).

The first trial of SAS, then call The Junior Detective Program, (Beaumont & Sofronoff, 2008) was a randomized controlled trial that demonstrated significant positive outcomes for children in both social and emotional domains as reported by parents and demonstrated by the child's capacity to generate meaningful strategies to cope with difficult social situations. Following the initial trial, future trials focused on using the program in more diverse ways. This included a briefer parent-directed trial (Sofronoff, Silva & Beaumont, 2015), a school-based trial using different doses of intervention (Beaumont, Rotolone & Sofronoff, 2015), and a larger scale trial within several schools in NSW, Australia (Einfeld et al., 2018). These trials used the small group program and worked only with those children on the autism spectrum. In each of these trials positive outcomes were reported based on either parent or teacher report, as well as increases in child ability to generate strategies to cope with anxiety or anger provoking situations. The program was further trialed in a clinic setting with a typically developing population of children who were reported to have either emotional difficulties or peer relationship difficulties and similar results were found (Beaumont, Pearson & Sofronoff, under review). A randomized controlled trial of the brief parent-directed version of the program was also conducted (Beaumont, Walker, Weiss, & Sofronoff, under review) with similar results reported by parents.

While the SAS small group program has a considerable and growing evidence base for its use in both the original form and variant lighter touch formats, the decision to use the SAS program as the basis for the whole of classroom trial was only undertaken after careful consideration of alternative programs available and the utility of such programs in the context of a mainstream classroom environment.

1.2 Systematic Review

A systematic review was undertaken to explore the efficacy of social-emotional programs for children with autism post 2010 and to consider the extent to which these are offered within

mainstream classroom settings. While the current trial targeted primary-school aged children in year 5, a broader age bracket (school-aged children) was employed in the systematic review in order to obtain a large enough sample to allow for meaningful analysis. Thus, the systematic review examined the availability, quality and outcome of social, emotional, and social-emotional interventions available to people of all school age who had a diagnosis of autism. The question as to whether comorbidity was considered within study design was included as an additional focus of the review. The original Secret Agent Society trial of 2008 was excluded from this review but is included in an earlier review paper (Cappadocia & Weiss, 2011).

1.2.1 Method

The design, conduct, and reporting for the current systematic review was guided by the Centre for Reviews and Dissemination manual (CRD: Chambers et al., 2009). Based on the CRD, an initial review protocol including definition of terms, search strategy and inclusion criterion was drafted prior to the identification and selection of studies.

1.2.1.1 Definition of terms

Prior to conducting a literature search, key terms were defined and outlined to aid in the selection of studies. For the purpose of this review a social skill intervention is defined as a program designed to improve the social function of an individual, across verbal, nonverbal and behavioural domains (White, 2007; Frankel et al., 2010). An emotion skill intervention is defined as a program designed to specifically enhance or improve a child's emotion recognition and emotion regulation skills, where emotion recognition is taken to mean the ability to identify and understand the meaning of specific emotions, and emotion regulation is taken to mean the ability to modulate one's own emotional experience (Beaumont & Sofronoff, 2008; Weiss, 2014). Finally, a social-emotional intervention is defined as a program that combines elements of both social and emotional intervention (Beaumont et al., 2015). The term 'youths' is taken to capture children and adolescents of school age (6 – 18 years) (Lawrence et al., 2015). This age range was chosen as current research suggests this is a critical period for intervention delivery for those on the autism spectrum (Picci & Scherf, 2015; White et al., 2007).

1.2.1.2 Search strategy

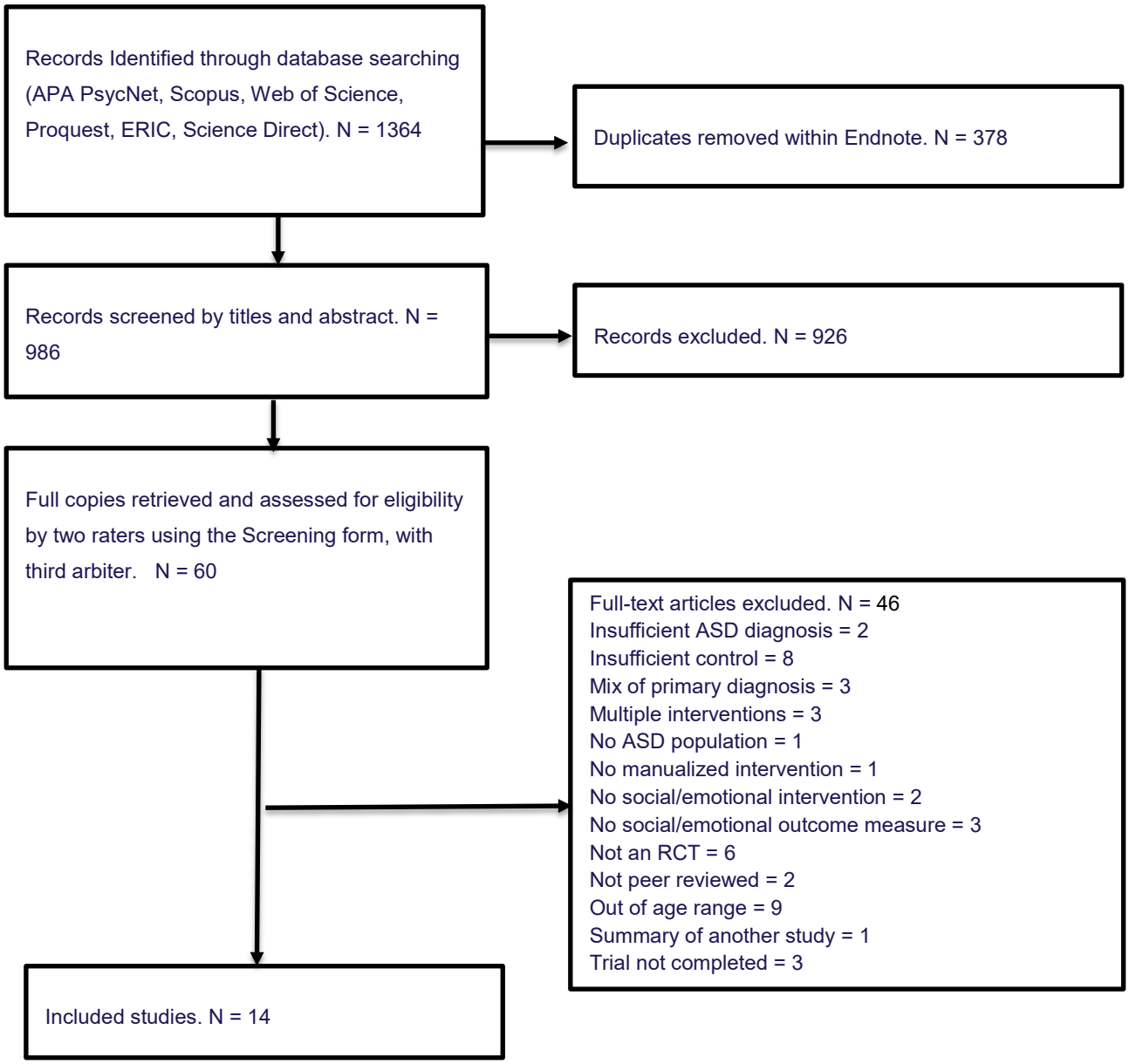
The APA PyscNet (PsycInfo and PsycArticles), Scopus, Web of Science, ProQuest, ERIC and Science Direct databases were searched in June 2016. The search strategy was developed in consultation with a librarian specializing in postgraduate psychology research at the University of Queensland and was run identically across each database. The final search strategy was

designed to capture three components: autism spectrum disorder, intervention type (social, emotional or social-emotional), and intervention design (program, course, intervention etc.). While truncation (e.g. multiple versions of a search term such as emotion, emotions, emotional, etc.) is often used for systematic review it was not used in this instance as the combination of both phrase search and truncation search was not possible on the APA PsycNet database at the time. The search was completed using Boolean String sequences with the operators AND/OR, “exact phrase” and “all fields” search techniques as this method could be conducted across each database. Appendix A provides a detailed description of the final search strategy.

1.2.1.3 First stage of the literature search

In the first stage of the review, the database search yielded 1,364 results, which were imported into EndNote Version 7 for Mac OS (Reuters, 2013). After removing duplicated search returns, 936 results remained. In the first stage of the review, titles and abstracts of these results were scanned and assessed for eligibility against the following inclusion criterion: a) participants aged 6 – 18 years; b) participants had a diagnosis of ASD or Asperger syndrome made by a medical health professional, meeting the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) or DSM-V standards (American Psychiatric Association, 2013); c) participants’ ASD diagnosis had been confirmed and screened using a validated measure for the current study; d) participants with a comorbid psychopathology were not excluded; e) the intervention was manualized and replicable; f) intervention specified as social, emotional or social-emotional in design; g) intervention was delivered to an individual or group in any setting (e.g. clinic, school, home); h) studies contained one intervention and one control; i) study design specified either randomized controlled trials or active control trials with an established treatment as usual (TAU) for the control; j) study published in English; k) study was published after 2010; l) study measured program efficacy for social, emotional, social-emotional outcomes pre and post intervention. After this process, 60 results remained. Figure 1 illustrates the process of study selection.

Figure 1. Flowchart of study selection process



1.2.1.4 Second stage of the literature search

In the second stage of the review, full text versions of the 60 articles were retrieved and assessed for eligibility using a screening form based on the process detailed by the CRD guidelines (Chambers et al., 2009). Two raters (both PhD students) independently read and assessed each article. Arbitration of disagreements as to whether to include or exclude studies between the independent raters was performed by a third reviewer (an Associate Professor with over 10 years clinical experience). From the pool of 60 papers, studies were excluded due to study design (insufficient control, N = 8), population (out of age range, N = 9; mix of primary

diagnoses in sample, N = 3; insufficient ASD diagnosis, N = 2; No ASD in sample, N = 1), lack of social or emotional focus (no social or emotional outcome measure, N = 3; no social or emotional intervention, N = 2), and methodological issues (not peer reviewed, N = 2; trial not completed, N = 3; data from another study, N = 2; summary of another study, N = 1). At the end of this process, 14 results remained in the review.

1.2.1.5 Data extraction

Data from the 14 results was extracted using the systematic process detailed by the CRD guidelines (Chambers et al., 2009). Data compiled in the extraction form included general information such as: inclusion criterion, participant characteristics, program outcomes, quality of study design, and type of statistical analysis used. Data regarding presence of comorbidity was also assessed. This included type and prevalence of comorbidity, how it was diagnosed, and whether it was addressed in intervention or measured in outcome.

1.2.1.6 Quality of study design

A quality assessment was embedded within the data extraction form used for this review. The aim was to capture any bias and errors in study design or methodology that can result in inaccurate measures of intervention outcome (Chambers et al., 2009). Integral to the assessment of study quality is the inclusion of an intention to treat and power analysis (Chambers et al., 2009; Whear et al., 2010; Lubans et al., 2012). An intention to treat analysis can be used to mitigate the risk of bias on intervention outcome that can arise from withdrawals, attrition and exclusions that occur during the study. It also provides a more conservative estimate of intervention efficacy (Chambers et al., 2009; Montori & Guyatt, 2001). A power analysis determines whether a study is able to detect an effect given its sample size (Chambers et al., 2009; Tabachnick & Fidell, 2013). Methodological quality was assessed by recording outcome measures, study design and the use of appropriate statistical methods (Chambers et al., 2009; Wise, 2004). Recording the presence of design specifics necessary for study replication, e.g. participant characterization, use of a manualized intervention, and use of a control group was also used to assess methodological quality (Chambers et al., 2009; White et al., 2007).

1.2.2 Results of systematic review

This review identified 14 studies that investigated social, emotional and social-emotional interventions for youths one the autism spectrum published since 2010. Seven were identified as primarily social skill interventions (see Table 1), and seven were identified as social-emotional skill interventions (see Table 2). There were no studies that focused solely on emotional intervention and outcome.

1.2.2.1 Social skill interventions

The studies that investigated the social skill interventions (N=7) included a total of 366 participants (28 to 76 participants per study) ranging in age from 7 to 18 years old. The interventions ran for a minimum of 5 to a maximum of 14 weeks, with all studies delivering intervention material in a group format. All seven studies required some level of parent involvement; with four studies (Frankel et al., 2010; Laugeson et al., 2012; Schohl et al., 2014; Yoo et al., 2014) delivering concurrent parent sessions that covered similar material to the child participant sessions. Two studies (Lopata et al., 2010; Thomeer et al., 2012) provided weekly parent sessions, and one study (Laugeson et al., 2014) provided weekly material handouts for parents.

Several studies investigated the same social skill intervention. An intervention entitled the Program for the Education and Enrichment of Relational Skills (PEERS) (Laugeson & Frankel, 2010) was used by four of the seven studies (Laugeson et al., 2012; Laugeson et al., 2014; Schohl et al., 2014; Yoo et al., 2014). The Skillstreaming program (McGinnis & Goldstein, 1997) was investigated by two studies (Lopata et al., 2010; Thomeer et al., 2012). The final study (Frankel et al., 2010) investigated the efficacy of an intervention entitled Children's Friendship Training (Frankel & Myatt, 2003).

1.2.2.2 Social skill intervention outcome measures

Each study used an assessment battery comprising several measures designed to assess the social and emotional outcomes of the intervention. Several measures were used across studies, for example the Social Responsiveness Scale (Constantino & Gruber, 2005) was used in four studies (Laugeson et al., 2014; Lopata et al., 2010; Thomeer et al., 2012; Yoo et al., 2014). Additionally, four studies (Frankel et al., 2010; Lopata et al., 2010; Schohl et al., 2014; Yoo et al., 2014) used the Quality of Play Questionnaire (Frankel & Mintz, 2011), three studies (Laugeson et al., 2014; Schohl et al., 2014; Yoo et al., 2014) used the Test of Adolescent Social

Skills Knowledge (Laugeson & Frankel, 2010), three studies (Frankel et al., 2010; Schohl et al., 2014; Yoo et al., 2014) used the Social Skills Rating System (Gresham & Elliott, 1990), and two studies (Lopata et al., 2010; Thomeer et al., 2012) used the Behavior Assessment System for Children, 2nd Edition (Reynolds & Kamphaus, 2004). Table 1 provides a comprehensive outline of each intervention and reported outcome.

1.2.2.3 Social-emotional skill interventions

Studies evaluating social-emotional skill interventions include a total of 230 participants (15 to 50 per study) ranging in age from 7 to 17 years old. The interventions ran for a minimum of 5 to a maximum of 24 weeks. The interventions were delivered in varying and mixed formats: three delivered intervention to groups of participants (Andrews et al., 2013; Koning et al., 2013; White et al., 2013), three provided intervention in a mix of individual and group sessions with family members (Storch et al., 2013; Storch et al., 2015; Wood et al., 2015), and five studies delivered intervention with individual sessions only (Storch et al., 2013; Storch et al., 2015; Thomeer et al., 2015; White et al., 2013; Wood et al., 2015).

Of the social-emotional programs included in this review, five investigated CBT based interventions. Three studies (Storch et al., 2013; Storch et al., 2015; Wood et al., 2015) investigated the efficacy of the Behavioral Intervention for Anxiety in Children with Autism (BIACA)(Wood & Drahota, 2005). Two studies investigated the efficacy of unnamed CBT based interventions (Andrews et al., 2013; Koning et al., 2013). Of the final two studies exploring social-emotional interventions, one (White et al., 2013) assessed the Multimodal Anxiety and Social Skills Intervention (MASSI) (White et al., 2010) and the other (Thomeer et al., 2015) explored a Mind Reading intervention (Baron-Cohen et al., 2004).

Table 1: Social skill interventions and outcomes

Description of intervention and reported outcome.

Author	Sample; age range, design and intervention	Social and emotional outcomes (delivery mode)	Analysis	Results
Frankel et al. (2010)	76; 7 – 10 RCT; delayed treatment control group *Children's Friendship Training* 60mins, 1 x week for 12 weeks Groups of 10 children (minimum 4 ASD) Parents and participants seen in concurrent separate sessions	The Loneliness Scale (TLS) (Asher et al., 1984) (Child) Piers-Harris Self-Concept Scale (PHS) (Piers, 1984) (Child) Social Skills Rating System (SSRS) (Gresham & Elliott, 1990) (Parent and teacher) Quality of Play Questionnaire (QPQ) (Frankel & Mintz, 2011) (Parent)	ANCOVA and pairwise t tests Follow up: 3-month; Intervention	Significant improvement on TLS pretest to posttest for Intervention compared to Control ($p < .025$). Significant improvement on the PHS mean popularity subscale pretest to posttest compared to Control ($p < .025$). Significant improvement on the SSRS (parent) self-control subscale pretest to posttest compared to Control ($p < .05$). SSRS (teacher) scales, ns. Significant improvement on the QPQ hosted play dates subscale pretest to posttest for Intervention compared to Control ($p < .0001$). Significant decrease in QPQ disengaged behaviors on play dates subscale pretest to posttest compared to Control ($p < .0001$). **Follow up T1 – T3**: Gains not maintained for child measures. Gains not maintained for teacher measures. Gains maintained for parent measures, with the exception of QPQ disengaged behaviors on play dates subscale, which had a significant reversal ($p < .0001$). **Follow up T2 – T3:** Gains not maintained for child measures. Gains not maintained for teacher measures. Intervention reported significant improvement on the SSRS (parent) subscales of assertion ($p < .0001$), self-control ($p < .005$), and internalizing behaviors ($p < .025$). Intervention reported significant increase on QPQ hosted play dates subscale ($p < .05$). Intervention reported significant decrease on QPQ disengaged behaviors on play dates subscale ($p < .025$) and conflict subscale ($p < .0005$).
Laugeson et al. (2012)	28; 12 – 17 RCT; delayed treatment control *PEERS*	Social Skills Rating System (SSRS) (Gresham & Elliott, 1990) (Parent and teacher)	MANOVA and Paired samples t tests Follow up: 14 week; Intervention only	Significant improvement on the SSRS total (parent) pretest to posttest for Intervention compared to Control ($p < .01$). Specifically, SSRS Cooperation subscale ($p < .01$), Assertion subscale ($p < .01$), Responsibility subscale ($p < .02$), and Self-Control subscale ($p < .02$). SSRS (teacher), ns.

	90mins, 1 x per week for 14 weeks Group therapy, participants and parents attend separate sessions Groups of 8 – 10 participants	Social Responsiveness Scale (SRS) (Constantino & Gruber, 2005) (Parent and teacher) Quality of Play Questionnaire (QPQ) (Frankel & Mintz, 2010) (Parent and child) Test of Adolescent Social Skills Knowledge Revised (TASSK-R) (Laugeson & Frankel, 2010) (Child)		Significant improvement in the SRS total (parent) pretest to posttest for Intervention compared to Control ($p < .02$). Specifically, Social Awareness subscale ($p < .02$), Social Cognition subscale ($p < .02$), Social Communication subscale ($p < .01$), Social Motivation subscale ($p < .05$) and Decreased Autistic Mannerisms subscale ($p < .05$). SRS (teacher) scales, ns. Significant increase on the QPQ (parent) hosted get-togethers subscale pretest to posttest for Intervention compared to Control ($p < .01$). Significant increase on to QPQ (adolescent) hosted get-togethers subscale on the pretest to posttest for Intervention compared to Control ($p < .03$). Significant increase on TASSK-R pretest to posttest for Intervention compared to Control ($p < .01$). **Follow up T1 – T3**: Treatment gains maintained at follow up with the exception of SRS-P social cognition. Additionally, Intervention reported significant decrease in Problem Behaviors ($p < .04$), and Externalizing Behaviors ($p < .01$) on the SSRS (parent). Intervention reported significant increase on SSRS (teacher) on Social Skills scale ($p < .03$) and Assertion scale (p < .02).
Laugeson et al. (2014)	73; 12 – 14 Active control; TAU (Super Skills, Coucouvanis et al., 2005) *PEERS* 30mins, 5 x per week for 14 weeks School based version: teacher facilitate Parent and participant involvement	Test of Adolescent Social Skills Knowledge (TASSK)(Laugeson & Frankel, 2010) (Child) Quality of Play Questionnaire (QPQ) (Frankel & Mintz, 2011) (Parent and child) Social Responsiveness Scale (SRS)(Constantino & Gruber, 2005) (Parent and teacher) Social Skills Rating System (SSRS)(Gresham & Elliott, 1990) (Parent and teacher)	GLM Follow up: none reported	Significant improvement on the TASSK pretest to posttest for Intervention compared to TAU ($p < .001$, $d = 1.88$). Significant improvement on the QPQ hosted ($p < .01$, $d = .82$), and invited ($p < .02$, $d = 0.59$) get-togethers, pretest to posttest for Intervention compared to TAU. Significant improvement on the SRS (Teacher) Social Responsiveness subscale ($p < .01$, $d = -.63$), Social Awareness subscale ($p < .03$, $d = - .59$), Social Communication subscale ($p < .03$, $d = -.57$), Social Motivation subscale ($p < .03$, $d = -.52$) and Decreased Autistic Mannerisms subscale ($p < .02$, $d = -.59$) from pretest to posttest for Intervention compared to TAU. SRS (Teacher) Social Cognition scale, ns. All parent measures, ns. SSRS, ns.

		Social Anxiety Scale (SAS)(La Greca & Lopez, 1998) (Parent and child) Friendship Qualities Scale (FQS)(Bukowski et al., 1994) (Child) Piers-Harris Self-Concept Scale (PHS) (Piers, 1984) (Child)		SAS, ns. FQS, ns. PHS, ns.
Lopata et al. (2010)	36; 7 – 12 RCT; waitlist control *Skillstreaming* 70mins, 5 x per day per week for 5 weeks Groups of 6 participants and 3 staff. Parents participate in weekly training groups (90mins).	Social Responsiveness Scale (SRS)(Constantino & Gruber, 2005) (Parent and staff) Behavior Assessment System for Children, 2nd Edition (BASC-2) (Reynolds, 2004) (Parent and staff) Diagnostic Analysis of Nonverbal Accuracy 2 (DANVA 2) (Nowicki, 1998) (Child)	ANCOVA and repeated measures ANOVA Follow up: none reported	Significant improvement on the SRS (Parent) pretest to posttest for Intervention compared to Control (p = .003, d = .625). Significant improvement on the SRS (Staff) pretest to posttest compared to Control (p = .003, d = .711). Significant improvement on the BASC-2 (Parent) Withdrawal scale pretest to posttest for Intervention compared to Control (p < .001, d = 1.055). BASC-2 (Parent) Social Skills scale, ns. Significant improvement on the BASC-2 (Staff) Withdrawal scale pretest to posttest for Intervention compared to Control (p = .007, d = .693). DANVA-2 (Child), ns.
Schohl et al. (2014)	63; 11 – 16 RCT; waitlist control *PEERS* 90mins, 1 x per week for 14 weeks Participant groups of maximum 10. Parents attend concurrent group session.	Test of Adolescent Social Skills Knowledge (TASSK)(Laugeson & Frankel, 2010) (Child) Quality of Socialization Questionnaire (QSQ) (Frankel & Mintz, 2008) (Parent and child) Friendship Qualities Scale (FQS)(Bukowski et al., 1994) (Child)	MANOVA with post hoc paired t tests ANOVA (teacher outcomes only) Follow up: none reported	Significant Time X Group interaction on TASSK (p <. 001, partial η^2 = 0.72). Intervention scoring significantly higher at posttest (p < .001). Significant Time x Group interaction for QSQ Hosted subscale (p <. 005, partial η^2 = 0.15) and Invited get-togethers subscale (p < .01, partial η^2 = 0.12). Intervention scoring significantly higher at posted for Hosted (p < .001) and Invited get-togethers at posttest (p <. 005). FQS, ns.

		Social Interaction Anxiety Scale (SIAS)(Mattick & Clarke, 1998) (Child) Social Responsiveness Scale (SRS)(Constantino & Gruber, 2005) (Parent and teacher) Social Skills Rating System (SSRS)(Gresham & Elliott, 1990) (Parent and teacher)		Significant Time x Group interaction for SIAS (p < .001, partial η^2 = 0.12). Intervention showing a significant decrease in social anxiety at posttest (p < .005). Significant Time x Group interaction for SRS (parent) (p < .01, partial η^2 = 0.14). Intervention scoring significantly lower at posttest (p < .001). SRS (teacher), ns. Significant Time x Group interaction for SSRS problem behaviours subscale (p < .05, partial η^2 = 0.06). Intervention showing a significant decrease at posttest (p < .005). Significant Group x Time interaction for SSRS (teacher) problem behaviours subscale (p < .05). Groups were comparable at pretest, with Intervention reporting significant decrease at posttest (p < .001). SSRS (teacher) social skills subscale, ns.
Thomeer et al. (2012)	35; 7 – 12 RCT; waitlist control *Skillstreaming* 70mins, 5 x per day per week for 5 weeks Groups of 6 participants and 3 staff. Parents participate in weekly training groups (90mins).	Social Responsiveness Scale (SRS)(Constantino & Gruber, 2005) (Parent and staff) Behavior Assessment System for Children, 2nd Edition (BASC-2) (Reynolds & Kamphaus, 2004) (Parent and staff) Diagnostic Analysis of Nonverbal Accuracy 2 (DANVA 2) (Nowicki, 1998) (Child)	ANCOVA and ANOVA Follow up: ANOVA Follow up: 2 – 3 month; Intervention (parent measures only)	Significant improvement on the SRS (Parent) pretest to posttest for Intervention compared to Control (p = .007, d = .67). Significant improvements on the SRS (Staff) pretest to posttest for Intervention compared to Control (p = .001, d = .86). Significant improvement on the BASC-2 (Parent) Social Skills scale pretest to posttest for Intervention compared to control (p = .011, d = .70). BASC-2 (Parent) Withdrawal, ns. Significant improvement on the BASC-2 (Staff) Withdrawal scale pretest to posttest for Intervention compared to Control (p = .012, d = .78). Significant improvement on the BASC-2 (Staff) Social Skills scale pretest to posttest for Intervention compared to Control (p = .003, d = .84). DANVA-2 (Child), ns. Follow up (T1 – T3): SRS (Parent), ns. Treatment gains maintained on the BASC-2 (Parent) Social Skills (p = .004, d = .68). BASC-2 (Parent) Withdrawal, ns.
Yoo et al. (2014)	55; 12 – 18 RCT; waitlist control *PEERS (Korean version)*	Korean Vineland Adaptive Behavior Scale (EHWA-VABS) (Sparrow & Cicchetti, 1985) (Parent) Test of Adolescent Social Skills Knowledge Revised	ANOVA Follow up: 3 month follow up; Intervention and Control	Significant Time x Group interaction on EHWA-VABS Interpersonal subscale (p <. 01). Intervention scoring significantly higher at posttest. Significant Time x Group interaction on TASSK-R (p < .01). Intervention scoring significantly higher at posttest.

	90mins, 1 x per week for 14 weeks Participant groups. Parents attend concurrent group session.	(TASSK-R) (Laugeson & Frankel, 2010) (Child) Quality of Play Questionnaire (QPQ) (Frankel & Mintz, 2010) (Parent and child) Korean Social Skills Rating System (K-SSRS)(Gresham & Elliott, 1990) (Parent and teacher) Social Communication Questionnaire (SCQ) (Lord & Rutter, 2003) (Parent) Social Responsiveness Scale (SRS)(Constantino & Gruber, 2005) (Parent and teacher) Child Depression Inventory (CDI) (Kovacs, 1985) (Child) State and Trait Anxiety Inventory for Children (STAIC) (Spielberger & Edwards, 1973) (Adolescent) Korean Child Behaviour Checklist (K-CBCL) (Achenbach, 1991) (Parent)	QPQ (parent and adolescent), ns. K-SSRS (parent and teacher), ns. SCQ (parent), ns. SRS (parent and teacher), ns. Significant Time x Group interaction on CDI ($p < .04$). Intervention scoring significantly higher at posttest. Significant Time x Group interaction on STAIC ($p <. 01$). Intervention scoring significantly higher at posttest. Significant Time x Group interaction on K-CBCL Anxiety/Depression subscales ($p < .03$), and the Internalising Problems subscale ($p < .02$). Intervention scoring significantly higher at posttest. Follow up (T2 - T3): EHWA-VABS, not measured. Treatment gains maintained on other measures. SCQ (parent), ns. SRS (parent and teacher), ns.

1.2.2.4 Social-emotional skill intervention outcome measures

All studies used an assessment battery comprising measures designed to assess the social and emotional outcome of the intervention. Within the assessment battery, each study used standard outcome measures, and reported significant findings pre- to post-test that was in favour of the intervention group on at least one outcome measure. Several measures were used across studies, for example the Social Responsiveness Scale (Constantino & Gruber, 2005) was used by six studies (Koning et al., 2010; Storch et al., 2013; Storch et al., 2015; Thomeer et al., 2015; White et al 2013; Wood et al., 2015). Additionally, four studies (Storch et al., 2013; Storch et al., 2015; White et al., 2013; Wood et al., 2015) used the Pediatric Anxiety Rating Scale (RUPP, 2002) and the Clinical Global Impressions-Improvement and Severity Scales (Guy, 1976). Three studies investigating the BIACA intervention (Storch et al., 2013; Storch et al., 2015; Wood et al., 2015) utilised a similar outcome battery, including the Multidimensional Anxiety Scale for Children (March et al., 1997), and the Anxiety Disorders Interview Schedule DSM-IV (Silverman & Albano, 1996). Two studies (Storch et al., 2013; Storch et al., 2015) used the Child Behavior Checklist (Achenbach et al., 2001) and the Columbia Impairment Scale (Bird et al., 1993), and two studies (Storch et al., 2015; Wood et al., 2015) used the Revised Child Anxiety and Depression Scale (Chorpita et al., 2000). Table 2 provides a comprehensive outline of each intervention and reported outcome.

1.2.2.5 Assessment of methodological quality

Assessment of the methodological quality of the studies within the review highlighted several strengths that were shared across studies. All studies provided sufficient design specifics that would allow for study replication (White et al., 2007). All studies used appropriate randomization techniques to allocate participants to control and intervention conditions, and standard social and emotional outcome measures to determine the effectiveness of intervention (Chambers et al., 2009). However, only two studies reported an intention to treat analysis (White et al., 2013; Wood et al., 2015) and only seven reported conducting a power analysis (see Table 3).

Table 2: Social-emotional interventions

Description of intervention and reported outcome.

Author	Sample; age range, design and intervention	Social and emotional outcomes (delivery mode)	Analysis	Results
Andrews et al. (2013)	50; 7 – 12	Social Competence with Peers Questionnaire -Parent (SCPQ-P) (Spence, 1995) (Parent)	MANOVA and Independent samples t tests	SCPQ-P, ns.
	RCT; waitlist control *CBT Intervention* 120mins, 1 x per week for 5 weeks	Affection for Others Questionnaire for children with Asperger's syndrome (AOQ) (Sofronoff et al., 2014) (Parent)	Follow up: 3 month, Intervention only	Significant Time x Group interaction on AOQ total (p = .002, partial η^2 = .158). Intervention scoring significantly higher at posttest (p < .05). Significant effect of time on AOQ subscales (p = .005, partial η^2 = .213). Significant interaction between time and group on AOQ subscales (p = .015, partial η^2 = .174). Specifically, significant Time x Group interaction on AOQ Giving Affection subscale (p < .001, partial η^2= .171), with Intervention scoring significantly higher at posttest (p = .028). AOQ Receiving Affection subscale, ns. AOQ Communicating Empathy subscale, ns.
	Groups of 3 or 4 participants	General Affection Questionnaire (GAQ) (Sofronoff et al., 2014) (Parent)		GAQ, ns.
		Walk in the Forest Test (WFT) (Attwood, 2004) (Child)		WFT, ns. SCAS-P, ns.
		Spence Children's Anxiety Scale Parent version (SCAS-P) (Nauta et al., 2004) (Parent)		Follow-up T2 – T3: AOQ total, ns - treatment gains maintained. Within the AOQ subscales, significant main effect of time (p = .026, partial η^2 = .38). Significant interaction between time and group,

				where AOQ Communicating Empathy subscale increased between posttest and follow up (p = .002, partial η^2= .37).
Koning et al. (2013)	15; 10 – 12 RCT *CBT Based Social Skills Intervention* 120 mins, 1 x per week for 15 weeks Groups of 7 participants with 'leader'	Vineland Adaptive Behavior Scales – Second Edition (VABS-2) (Sparrow et al., 2005) (Parent) Social Responsiveness Scale (SRS) (Constantino & Gruber, 2005) (Parent) Child and Adolescent Social Perception Measure (CASP) (Magill-Evans et al., 1995) (Child) Peer Interaction Measure (PIM) (Koning et al., 2008) (Independent raters) Test of Social Knowledge (TSK) (Koning et al., 2013) (Child)	ANOVA Follow up: none reported	VABS-2, ns. SRS, ns. Significant Time x Group interaction on CASP total (p < .003). Groups were comparable at pretest, with Intervention scoring significantly higher at posttest on the CASP Emotion Recognition subscale (p < .001, partial η^2= .61), and the Nonverbal Cues subscale (p < .006, partial η^2= .45). Significant Time x Group interaction on PIM, (p < .046, partial η^2= .29). Groups differed at baseline, with Intervention scores increasing at posttest compared to Control. Significant Time x Group interaction on TSK (p < .001, partial η^2 = .68). Groups differed at baseline, with Intervention scores increasing at posttest compared to control.
Storch et al. (2013)	45; 7 – 11	Pediatric Anxiety Rating Scale (PARS) (RUPP, 2002) (Clinician)	ANCOVA	Significant reduction on PARS pretest to posttest for Intervention compared to TAU (p < .01, d = 1.03).

RCT with TAU TAU: designed to reflect typical community treatment services. *BIACA CBT* 90mins, 1 x per week for 16 weeks Participant and family, individual and joint sessions	Anxiety Disorders Interview Schedule DSM-IV (ADIS) (Silverman & Albano, 1996) (Clinician) Clinical Global Impressions-Improvement and Severity Scale (CGI-S) (Guy, 1976) (Clinician) Multidimensional Anxiety Scale for Children (MASC) (March et al., 1997) (Parent) Social Responsiveness Scale (SRS) (Constantino & Gruber, 2005) (Parent) Child Behaviour Checklist (CBCL) (Achenbach & Rescorla, 2001) (Parent)	Follow up: 3 month, Intervention only	Significant reduction on ADIS pretest to posttest for Intervention compared to TAU ($p < .01$, $d = 0.84$). Significant reduction on CGI-S pretest to posttest for Intervention compared to TAU ($p < .05$, $d = 1.06$). MASC, ns. Significant improvement on SRS total (parent) pretest to posttest for Intervention compared to TAU ($p < .001$, $d = .73$). Specifically, the Social Communication subscale ($p < .05$, $d = 0.53$) and Decreased Autistic Mannerisms subscale ($p < .05$, $d = 0.50$). Significant improvement on the CBCL Internalizing Behaviors pretest to posttest for Intervention compared to TAU ($p <. 05$, $d = 0.56$). CBCL Externalizing Behaviours, ns. Significant improvement on the CIS pretest to posttest for Intervention compared to TAU ($p < .01$, $d = 0.73$).

		Columbia Impairment Scale (CIS) (Bird et al., 1993) (Parent) Revised Children's Manifest Anxiety Scale (RCMAS) (Reynolds & Richmond, 1985) (Child)		Significant improvement on the RCMAS Anxious Arousal subscale pretest to posttest for Intervention compared to Control ($p < .05$, $d = .63$). All other RCMAS subscales, ns. Follow up T1 – T3: A number of significant results were observed from pretest to follow up for the PARS ($p < .01$, $d = 1.42$), ADIS ($p < .01$, $d = .98$), CGI-S ($p < .01$, $d = .95$), MASC ($p < .01$, $d = .80$), SRS total ($p < .05$, $d = .73$), SRS Awareness ($p < .05$, $d = .64$), SRS Autistic Mannerisms ($p < .05$, $d = .60$), CBCL Internalizing ($p < .01$, $d = .83$), CBCL Externalizing ($p < .05$, $d = .73$), RCMAS Dysphoric Mood ($p < .05$, $d = .63$). Follow up T2 – T3: Treatment gains maintained, with the exception of RCMAS Dysphoric Mood subscale ($p < .05$) and Oversensitivity Subscale ($p < .05$); both of which were significantly lower than at posttest.
Storch et al. (2015)	25; 11 – 16 RCT with TAU TAU: community treatment services. *BIACA CBT*	Anxiety Disorders Interview Schedule DSM-IV (ADIS) (Silverman & Albano, 1996) (Clinician) Pediatric Anxiety Rating Scale (PARS) (RUPP, 2002) (Clinician)	ANCOVA and paired t tests Follow up: 1 month, Intervention only	Significant reduction on the ADIS pretest to posttest for Intervention compared to Control ($p < .001$, $d = 1.30$). Significant reduction on the PARS pretest to posttest for Intervention compared to Control ($p < .05$, $d = 0.79$). Significant reduction on the CGI-S pretest to posttest for Intervention compared to Control ($p < .05$, $d = 0.94$).

60 – 90 mins, 1 x per week for 16 weeks Participant and family, individual and joint sessions	Clinical Global Impressions-Improvement and Severity Scale (CGI-S)(Guy, 1976) (Clinician)	Significant reduction on the CIS-P pretest to posttest for Intervention compared to Control ($p < .05$, $d = 0.59$).
	Columbia Impairment Scale (CIS)(Bird et al., 1993) (Parent)	Significant reduction on the CBCL Externalizing Behaviours subscale pretest to posttest for Intervention compared to Control ($p < .001$, $d = 0.63$), compared to control.
	Child Behaviour Checklist (CBCL) (Achenbach & Rescorla, 2001) (Parent)	CBCL Internalizing Behaviours subscale, ns. MASC-P, ns.
	Multidimensional Anxiety Scale for Children (MASC-P) (March et al., 1997) (Parent)	Significant improvement on the SRS total pretest and posttest for Intervention compared to Control ($p < .001$, $d = 0.84$). Specifically, Social Awareness subscale ($p < .05$, $d = 0.75$), Social Cognition subscale ($p < .001$, $d = 0.76$), and Social Communication subscale ($p < .001$, $d = 1.13$).
	Social Responsiveness Scale (SRS) (Constantino & Gruber, 2005) (Parent)	RCADS, ns.
	Revised Child Anxiety and Depression Scale (RCADS)(Chorpita et al., 2000) (Child)	Follow up T2 – T3: Treatment gains maintained at follow up.

				Significant improvement noted on SRS total scales (p <.001, d = 1.89). Specific subscales of significance were social cognition (p < .001, d = 1.55), social communication (p < .001, d = 1.19), social motivation (p < .001, d = 1.13) and autistic mannerisms (p < .001, d = 1.39).
Thomeer et al. (2015)	43; 7 – 12 RCT; waitlist control *Mind Reading and In Vivo Rehearsal* (Baron Cohen et al., 2004) 90mins, 2 x per week for 24 sessions Individual sessions with staff clinician	Cambridge Mindreading Face-Voice Battery for Children (CAM-C)(Golan et al., 2006) (Child) Emotion Recognition and Display Survey (ERDS) (Thomeer et al., 2011) (Child) Social Responsiveness Scale (SRS) (Constantino & Gruber, 2005) (Parent) Behavior Assessment System for Children, Second Edition – Social Skills scale only (BASC-2) (Reynolds & Kamphaus, 2004) (Parent)	ANCOVA and t test Follow up: 5 week, Intervention and Control	Significant between-groups difference for CAM-C Faces (p < .001, ω^2 = .23). Intervention scoring significantly higher at posttest (p < .001, d = 1.34). Significant between-groups difference for CAM-C Voices (p < .001, ω^2= .14). Intervention scoring significantly higher at posttest (p < .001, d = 0.99). Significant between-groups difference for ERDS Receptive subscale (p = .006, ω2 = .08). Significant between-groups difference for ERDS Expressive subscale (p = .0025, ω^2 = .11). Intervention scoring significantly higher at posttest (p = .0125, d = 0.61). Significant between-groups difference on SRS total (p = .0135, ω^2 = .04). Intervention scoring significantly higher at posttest (p = .0175, d = .46). BASC-2, ns. Follow up T2 – T3 for both groups:

				Significant difference between groups CAM-C Faces, with Intervention scoring higher (p = .001, d = 0.86). Significant difference between groups CAM-C Voices, with Intervention scoring higher (p = .006, d = 0.66). Significant difference between groups on ERDS Receptive subscale, with Intervention scoring higher (p = .0045, d = .73). Significant difference between groups on ERDS Expressive subscale, with Intervention scoring higher (p = .003, d = .85). Significant difference between groups on SRS total (p = .023, d = .45)
White et al. (2013)	30; 12 – 17 RCT; waitlist control *Multimodal Anxiety and Social Skills Intervention Program (MASSI)* Individual therapy: 60 – 70mins, 1x week for 13 sessions	Social Responsiveness Scale (SRS) (Constantino & Gruber, 2005) (Parent) Child and Adolescent Symptom Inventory 4 ASD Anxiety Scale (CASI-Anx) (Sukhodolsky et al., 2008) (Parent) Pediatric Anxiety Rating Scale (PARS) (RUPP, 2002) (Blind clinician rater)	ANCOVA, t tests and Pearson Chi-Square test Follow up: none reported	Significant between-groups difference on SRS total (p = .007), with treatment showing significant improvement between baseline and post-intervention (p < .01, d = 1.18). CASI-Anx, p = .22. PARS, p = .31. CGI-I, ns. Significant between-groups difference on DD-CGAS (p = .029), with treatment showing significant improvement between baseline and post-intervention (p <. 01, d = 0.81).

	Group therapy: 75 mins, 1 x week for 7 sessions Parent education and coaching: 15 mins, 1 x week for 13 sessions	Clinical Global Impressions-Improvement Scale (CGI-I) (Guy, 1976) (Blind clinician rater) Developmentally Disabled Children's Global Assessment (DD-CGAS) (Wagner et al., 2007) (Blind clinician rater)		
Wood et al. (2015)	33; 11 – 15 RCT; waitlist control *Behavioural Interventions for Anxiety in Children with Autism (BIACA CBT)* 90mins, 1 x per week for 16 weeks Participant and family, individual and joint sessions	Anxiety Disorders Interview Schedule DSM-IV (ADIS) (Silverman & Albano, 1996) (Clinician) Clinical Global Impressions-Improvement and Severity Scale (CGI-I; CGI-S) (Guy, 1976) (Clinician) Multidimensional Anxiety Scale for Children (MASC-P) (March et al., 1997) (Parent)	ANCOVA and paired t tests Follow up: 1 month, Intervention	ADIS, ns. Significant difference in CGI-I positive treatment response criteria in treatment compared to Control ($p < .01$, OR = 9.38). MASC-P, ns. Significant between-groups difference on PARS, with Intervention reporting lower anxiety scores compared to Control ($p = .04$, $d = 0.74$). RCADS, ns.

	Pediatric Anxiety Rating Scale (PARS) (RUPP, 2002) (Clinician)	Significant between-groups difference on the SRS, with Intervention reporting improvement in ASD symptoms compared to Control ($p < .01$, $d = 1.17$).
	Revised Child Anxiety and Depression Scale (RCADS)(Chorpita et al., 2000) (Child)	Follow up T2-T3: Only analyzed 13/15 CGI-I positive responders.
	Social Responsiveness Scale (SRS) (Constantino & Gruber, 2005) (Parent)	Reduction in principal diagnosis on ADIS ($p = .02$, effect size not reported) MASC-P, ns. Maintenance of treatment effects indicated on PARS, with significant change at follow up ($p = .79$, $d = 0.07$). Significant reduction in anxiety from posttest to follow up on RCADS ($p = 0.2$, $d = 0.95$). SRS, ns.

Table 3. Intention to treat and/or power analysis

Methodology of study

Study	Intention to treat	Power Analysis
Andrews et al., 2013	X	X
Frankel et al., 2010	X	Yes
Koning et al., 2013	X	Yes
Laugeson et al., 2012	X	X
Laugeson et al., 2014	X	X
Lopata et al., 2010	X	Yes
Schohl et al., 2014	X	X
Storch et al., 2013	X	Yes
Storch et al., 2015	X	Yes
Thomeer et al., 2012	X	X
Thomeer et al., 2015	X	X
White et al., 2013	Yes	Yes
Wood et al., 2015	Yes	Yes
Yoo et al., 2014	X	X

1.2.2.6 Reporting of clinical significance

Aside from reporting statistically significant results, an additional and important aspect of methodological design when assessing intervention efficacy is the reporting of clinical significance (Wise, 2004). While all studies reported statistically significant findings on outcome measures, clinical significance as a measure of treatment outcome was only addressed in seven of the fourteen studies (Frankel et al., 2010; Koning et al., 2013; Lopata et al., 2010; Storch et al., 2013; Storch et al., 2015; White et al 2013; Wood et al., 2015).

Of the seven studies, Frankel and colleagues (2010), and White and colleagues (2013) used the Reliable Change Index (RCI) to ascertain clinical significance of outcome measures (Jacobson & Truax, 1991; Wise, 2004). Using this method, Frankel and colleagues (2010) reported that 91.4% of the Intervention group and 87.1% of the Delayed Treatment Control group showed reliable change on one outcome measure at posttest, with this trend continuing for 66.7% of the sample at follow up. White and colleagues (2013) reported RCI for their chosen primary outcomes, with 60% of the Intervention group and 0% of the Control group showing clinically significant improvement on the Social Responsiveness Scale (SRS; Constantino & Gruber, 2005), $\chi^2(1) = 12.86$, $p < .001$, and a non-significant result reported for the Child and Adolescent Symptom Inventory 4 ASD Anxiety Scale (CASI-Anx; Sukhodolsky et al., 2008).

The remaining five studies assessed clinical significance in response to treatment via recorded changes in clinical categories on outcome measures. For example, three studies (Storch et al., 2013; Storch et al., 2015; Wood et al., 2015) recorded changes in severity of anxiety symptoms on the CGI-S (CGI-S; Guy, 1976). The CGI-S measures severity of anxiety symptoms, that ranges from 0 (no symptoms) to 6 (extremely severe symptoms). In more detail, Storch and colleagues (2013) reported that on the CGI-S, 75% of the Intervention group reached non-clinical levels on the CGI-S compared to 14% of the Control group ($p < .01$, $d = 1.37$). Storch and colleagues (2015) reported that 68.8% of the Intervention group reached non-clinical levels compared to 26.7% of the Control group ($p = .03$). Wood and colleagues (2015) reported 79% of the Intervention group reached non-clinical levels compared to 28.6% of the Control group ($p > .01$, OR = 9.38). Additionally, the above three studies analyzed the effect of intervention on the severity of clinical symptoms using the Anxiety Disorders Interview Schedule DSM-IV (ADIS) (Silverman & Albano, 1996). Here, remission of anxious symptoms post intervention were quantified with Storch and colleagues (2013) reporting 38% of the Intervention group reached clinical remission posttest compared to 5% of the Control group ($p < .01$, $d = 1.37$), Storch et al., 2015 reported that 37.5% of Intervention reached clinical remission compared to 0% of the Control ($p = .02$), and Wood and colleagues (2015), reported ns results. Of the two final studies to report clinical impact, Lopata and colleagues (2010) used the SRS (SRS; Constantino & Gruber, 2005) as a measure of clinical change, and reported that 50% of the Intervention group had reduced social impairment by one clinical range (either from severe to mild-to-moderate, or mild-to-moderate to normal range). Koning and colleagues (2013) reported that the Intervention group improved by an average of 10 points on the Child and Adolescent Social Perception Measure (CASP; Magill-Evans et al., 1995) was clinically significant.

1.2.2.7 Follow up

Nine studies reported follow up of outcome at a post intervention time point, with the length of follow up ranging from 4 weeks to 12 weeks (see Table 4). Seven studies reported follow up for the intervention condition alone (Andrews et al., 2013; Frankel et al., 2010; Laugeson et al 2012; Storch et al., 2013; Storch et al., 2015; Thomeer et al., 2010; Wood et al., 2015), with two studies reporting follow up for both intervention and control groups (Thomeer et al., 2015; Yoo et al., 2014). Follow up outcomes were analysed by comparing time 1 (T1) to follow-up (T3) by two studies (Laugeson et al., 2012; Thomeer et al., 2012), while six reported a comparison of immediately post-intervention (T2) to T3 (Andrews et al., 2013; Frankel et al., 2010; Storch et al., 2015; Thomeer et al., 2015; Wood et al., 2015; Yoo et al., 2014), and one study (Storch et al., 2013) reported both. The intervention outcomes at follow up are displayed in Tables 1 and 2.

Table 4: Studies reporting follow up

Availability of follow up data and time period of data collection

Study	Follow up reported?	T1 – T3	T2 – T3	Length
Andrews et al., 2013	Yes	X	Yes	12 weeks
Frankel et al., 2010	Yes	X	Yes	12 weeks
Koning et al., 2013	X	X	X	X
Laugeson et al., 2012	Yes	Yes	X	14 weeks
Laugeson et al., 2014	X	X	X	X
Lopata et al., 2010	X	X	X	X
Schohl et al., 2014	X	X	X	X
Storch et al., 2013	Yes	Yes	Yes	12 weeks
Storch et al., 2015	Yes	X	Yes	4 weeks
Thomeer et al., 2015	Yes	X	Yes	5 weeks
Thomeer et al., 2012	Yes	Yes	X	8 – 12 weeks
White et al., 2013	X	X	X	X
Wood et al., 2015	Yes	X	Yes	4 weeks
Yoo et al., 2014	Yes	X	Yes	12 weeks

1.2.2.8 Comorbidity

Of the fourteen studies within the review, eight addressed comorbidity within the study design, two excluded participants with comorbidities, and four made no reference to comorbidity within their sample, study design or outcomes (see Table 5). Of the eight studies to address comorbidity, seven reported on comorbid presentation of autism and anxiety (Andrews et al., 2013; Schohl et al., 2014; Storch et al., 2013; Storch et al., 2015; White et al., 2013; Wood et al., 2015; Yoo et al., 2014), and one reported their sample contained several comorbid presentations of ASD and Attention-Deficit/Hyperactivity Disorder, Major Depressive Disorder, Adjustment Disorder and Generalized Anxiety Disorder (Laugeson et al., 2012). Four specified that participants must have a comorbid anxiety disorder to participate (Storch et al., 2013; Storch et al., 2015; White et al., 2013; Wood et al., 2015). In each study, an anxiety diagnosis was determined using standard measures with these results included as baseline participant characteristics. Anxiety was also a focus of intervention, with each study using standard outcome measures to monitor change in anxiety symptoms from pre to post intervention. Three of the four studies reported that intervention resulted in a significant reduction of anxiety symptoms as recorded by the assessment battery (Storch et al., 2013, Storch et al., 2015; Wood et al., 2015). Two studies (Storch et al., 2013; Storch et al., 2015) reported a clinically significant reduction in anxiety post intervention.

An additional three studies did not require participants to have a comorbid diagnosis but did state that anxious symptoms were a focus of intervention. Standard measures used to monitor change in anxiety symptoms pre to post intervention (Andrews et al., 2013; Schohl et al., 2014; Yoo et al., 2014). Two of the three studies reported that intervention resulted in a significant reduction of anxiety symptoms (Schohl et al., 2014; Yoo et al., 2014). The final of the eight studies to address comorbidity provided details of several comorbid presentations within participant characteristics, however did not follow up comorbid presentation with outcome measurement (Laugeson et al., 2012).

Table 5: Studies reporting comorbidity

Management of comorbidity within each study.

Address comorbidity	Do not address comorbidity	Actively exclude comorbidity
Andrews et al., 2013	Koning et al., 2013	Frankel et al., 2010
Laugeson et al., 2012	Lopata et al., 2010	Laugeson et al., 2014
Schohl et al., 2014	Thomeer et al., 2012	
Storch et al., 2013	Thomeer et al., 2015	
Storch et al., 2015		
White et al., 2013		
Wood et al., 2015		
Yoo et al., 2014		

1.3 Discussion

The aim of this review was to compile and assess the quality and outcome of current social, emotional, and social-emotional interventions on offer for youths one the autism spectrum. While no studies that focused solely on emotional intervention and outcome were identified, a number of social and social-emotional interventions were reviewed. Across the studies in this review eight interventions were investigated, with every study reporting significant improvement on a least one social and emotional outcome for participants.

Of social skill interventions available, the PEERS program was the most represented and generally produced positive outcomes. Three papers (Laugeson et al., 2012, Laugeson et al., 2014; Schohl et al., 2014) reported significant improvement in social function on outcome measures. One paper (Yoo et al., 2014) reported mixed results on program efficacy, however this could be due to the English – Korean translation of the intervention. The BIACA program was the most frequently used social-emotional intervention. The reported efficacy of the BIACA intervention was generally

positive, with two studies (Storch et al., 2013 and Storch et al., 2015) reporting significant improvement on social and emotional outcomes and one paper (Wood et al., 2015) reporting mixed results, where some social and emotional were not maintained at follow-up.

Comparing the overall efficacy of the social and social-emotional interventions between studies proved challenging. All studies within this review used varied definitions of 'social' and social outcome reporting also varied substantially between studies. For example, Wood and colleagues (2014) conceptualized social behavior in terms of social anxiety levels, whereas Attwood and colleagues (2013) conceptualized social behavior as level of affectionate behaviour. Conversely, Frankel and colleagues (2010) reported the use of a self-esteem measure (the PHS) to monitor social outcomes of intervention. A more unified conceptualization of what is "social" and what is "emotional" is required to enable intervention efficacy to be compared. Additionally, several social skill interventions (Frankel et al., 2010; Laugeson et al., 2014; Yoo et al., 2014; Lopata et al., 2010; Thomeer et al., 2012) report improvement on measures of emotional function. This highlights the presence of an emotional element within social exchanges, and the interplay between social and emotional behavior in ASD. Future studies of intervention efficacy would be strengthened by including measures of both social and emotional behavior irrespective of whether the intervention is labeled social or social-emotional.

When considering the design and methodology of the studies included in this review, there were some flaws and future areas of improvement. All studies collected outcome data from participants and parent and teacher responders who were not blind to treatment group, which may have led to biased responding on outcome. This is a challenge within the field, where demand characteristics may be influencing reported outcomes (Colman, 2008). Similarly, of the eleven studies that include parent report and participation, five did not provide demographic information on parent participants (Frankel et al 2010; Laugeson et al 2012; Laugeson et al., 2014; White et al., 2013; Wood et al., 2015). This may have an impact on outcome and results, for example socioeconomic status of the parents may impact the responses collected, and future studies may benefit from recording and analyzing this information (Casagrande & Ingersoll, 2017; Pickard & Ingersoll, 2016). Additionally, numerous studies relied on the use of parent and teacher data, and reported high attrition from parent participants, with parents dropping out of intervention before follow up could be conducted (Laugeson et al., 2014) and poor follow up from teacher participants, where the assessment battery was not completed sufficiently for analyses to run (Laugeson et al., 2012).

Reviewing the quality assessment highlighted that several studies included in this review did not complete an intention to treat or power analysis, and consequently many risked being

underpowered. Many studies used standard clinical measures as part of their assessment battery yet only half the studies in the review reported clinical significance of intervention outcome (Frankel et al., 2010; Koning et al., 2013; Lopata et al., 2010; Storch et al., 2013; Storch et al., 2015; White et al 2013; Wood et al., 2015). In a clinical context, it is important to establish meaningful clinical improvement as well as statistically significant improvement (Wise, 2004). A lack of reporting in many of the studies in this review may be in part due to the small sample size and underpowered nature of several of the studies, limiting their ability to report meaningful results. Of the studies that did report clinical significance in the results, there was also some discrepancy. For example, Frankel et al., 2010 does not specify which measure reaches clinical significance, instead stating "at least one", and Koning et al., 2013 reported clinically significant findings on the CASP (Magill-Evans et al., 1995) but does not report the statistical method by which this was achieved. To allow for better comparison between interventions within the field and aid in reporting clinically significant outcome, it would be of benefit to consider a unified assessment battery that includes standard measures that can produce clinical significant reporting such as the SRS (Constantino & Gruber, 2005) or BASC (Reynolds & Kamphaus, 2002).

The reporting of follow up outcomes is integral to quantifying longevity of intervention effect (Chambers et al., 2009; Gates, Kang & Lerner, 2017). However, in this review, several studies did not report follow up outcomes. Projects may not have had sufficient funding to sustain follow up outcomes after pre and post assessment. Of the papers that did report follow up outcomes the length of follow up varied considerably between studies and there was discrepancy between the type of follow up reported, with some studies reporting T1 – T3 results and others reporting T2 – T3. In the future it would be of benefit to specify a standard period of acceptable follow up. This includes what is to be considered an appropriate length and type of follow up and would allow future studies to adhere to a standard in their design (Chambers et al., 2009).

Positively, the review demonstrated that all currently available interventions for youths with ASD, whether social-emotional or social, produce some improvement. However, a key barrier to successful ASD intervention selection and application is a lack of uniformity across numerous aspects of the current body of literature.

In looking at the systematic review it is clear that the need to work with social skills is best done in the context of ameliorating emotional regulation issues that will likely arise in social contexts – notably, several of the recent programs reviewed concentrate solely on social skills. The combination of social and emotional skill training further provides an inclusive context for students not on the autism spectrum who struggle with emotional issues such as anxiety, emotion regulation

or anger. The SAS program has already been evaluated with this population (Beaumont et al., under review). It is also important that a whole of classroom program fits within the school curriculum and several of the reviewed programs are very long and/or delivered individually. Therefore, the decision was taken to modify the SAS small group program to accommodate the needs of schools delivering in a mainstream context.

2. SAS-WOC Development

2.1 Introduction

The Secret Agent Society – Whole of Classroom program (known as SAS-WOC) was developed following extensive consultation with teachers. The consultation process was undertaken by A/Prof Sofronoff and Dr Beaumont and involved teachers from 12 schools across Queensland, Australia. In these meetings, teachers were asked to discuss their concerns about the social-emotional issues of students in grade 5, including those on the autism spectrum, those with other social-emotional difficulties as well as students with minimal difficulties. Teachers were asked about social and/or emotional programs that were currently being offered within their school setting and the 'key features' that made/would make such programs more useful to them, such as length of sessions, length of program; capacity to complete questionnaires on 3-4 occasions during the trial; support from school leaders and parents and so on.

Teachers were also shown a clip from the computer game that is central to the SAS program as well as topics covered within the program. They were encouraged to ask questions and to offer suggestions. Notes were taken of all suggestions and concerns raised by teachers.

Following these meetings, it was determined that the SAS small group program did cover the major topics considered important by teachers, that the computer game would appeal to students in grade 5, and that the program should be configured to fit within one school term and with one lesson delivered each week. The program was to be streamlined so that content would fit within the time frame and the activity-based fun approach to teaching concepts was highlighted to reflect what the teachers suggested would work best with their students.

All modifications required to the original SAS small group program format to meet the specifications outlined by the consultation process were undertaken by Dr Renae Beaumont.

2.2 Developing the face-to-face (F2F) delivery model

The training materials and training protocol were modified to reflect the pared back SAS-WOC program, with a focus on a highly interactive F2F delivery undertaken over two consecutive days. The materials included participant notes containing all the training content and a full teacher manual containing all lesson plans for the program; all activities for students; and answers to any questions that might arise during delivery of the program.

The materials also included display posters for the classroom that depicted the major strategies taught and could be posted as those strategies were introduced. There was an additional poster that could be used to record incidents of students reporting that they or another student had used a target strategy. Each week had targets for students to achieve and success in achieving those targets earned rewards for the class. During training the teachers participated actively in each lesson and completed the activities in pairs or small groups as they would have their students do in class. The training was further supported by a PowerPoint slide presentation and video resources of students completing some of the activities.

The training was delivered by Dr Renae Beaumont with Dr Kate Sofronoff and clinical psychology students Monica McSweeney and Kathryn Farr offering support to the teachers as they completed their tasks.

2.3 Development of online delivery model

2.3.1 Rationale

Driven by the need to improve convenience, engagement and cost-effectiveness, an increase in online teacher professional development (PD) programmes has been witnessed in recent years (Bates, Phalen & Moran, 2016; Edinger, 2017). With rates of curriculum implementation post-PD teacher training having been shown to be equivalent whether conducted in a face-to-face or online environment (Fisher, Schumaker, Culbertson & Deschler, 2010; Fishman et al., 2016; Masters, de Kramer, O'Dwyer, Dash & Russell, 2012), the translation of SAS-WOC into an online interface was considered an essential step to maximising outreach and engagement with geographically diverse users across Australia in Year 2 of the trial. A 'hybrid' model of training, integrating a two-day online training program (core content) coupled with remote access to a content expert (via email, online forum, mobile) during training and throughout in-classroom program delivery was considered to offer the optimum combination of economy, convenience and support for educators in delivering the program.

2.3.1.1 Development of the SAS-WOC Online Teacher Training Course

The development of the training interface was divided into four main phases:

- Identification of an appropriate learning delivery platform
- Creation of multimedia resources to support online delivery
- Creation of training delivery structure
- Testing of the online interface

With a simple user interface and a broad range of built-in features, Moodle, a widely used, open-source learning platform was selected as an appropriate learning delivery platform for SAS-WOC online. Given the high teacher satisfaction ratings received from Year 1 'face-to-face' SAS-WOC training sessions, the mirroring of the in-person training components within the online interface was considered essential to replicating fidelity of instruction, as well as high levels of participant satisfaction (Schroeder, 2016). Accordingly, online delivery was split into six core modules as per the 'in person' training format (refer Table 6), while a range of novel, multimedia resources were created specifically to best promote personalised and (where applicable) collaborative learning (refer Table 7 for description and rationale). While SAS-WOC online was developed to function without synchronous content expert assistance, for the purpose of the evaluation trial, a course administrator/content expert was available (remotely) during all online training sessions to provide functional support and answer content enquiries. This additional support was included to encourage rapport and communication between the trial participants and research staff in order to facilitate direct and immediate feedback. Where appropriate, such feedback could then be employed to make amendments and/or recommendations for amendment in future iterations of the online training.

Table 6: SAS-WOC Program Online Teacher Training modules
Modules included in online training delivery

Module	Name
1	An Introduction to SAS-WOC
2	Evidence Base & Current Research
3	Content & Assessment Measures
4	Core Program Elements
5	Program Content & Delivery
6	Planning for Delivery

Table 7. Multimedia resources included in SAS-WOC Online

Description of resource included in SAS-WOC Online

Resource	Description	Key Points
Narrated Videos	MP4 video file created by combining PowerPoint slides, images, case studies and voice over.	• Employed best practice demonstration and modelling to provide key elements of successful PD (Schroeder, 2016; Zheng, 2015). • Video format allowed participants to pause or stop instruction at any time – offering great flexibility in delivery, considered a significant asset of online training (Collins & Liang, 2015).
Video Simulations	MP4 video file of specific case examples	• Video simulations operationalise key concepts and strategies to optimize teacher engagement (Schroeder, 2016; Zheng Lin, 2015). • For continuity in delivery, simulations were incorporated into the narrated video presentations.
Activity Handbook	Downloadable PDF	• Encourages participants to personalise training materials • Inclusion of an 'off-line' resource (PDF) was deliberately employed as a means of encouraging colleagues to work collaboratively when undertaking course activities. Encourages constructive dialogue (Prestridge &Tondeur, 2015) and the finding that teachers prefer to have access to both visual and written resources (Zheng, 2015). • Handbook contained interactive activities that teachers were asked to do during online training. Included discussing how they would apply the curriculum to the students in their class/school, role playing the delivery of core program elements with their colleagues (where possible) and problem-solving anticipated barriers to program delivery (e.g. program tailoring for students with attentional difficulties). Such activities create active and collaborative learning opportunities, as recommended by Desimone (2009).
Online Forum	Online chat room (multi-school participation)	• Monitored by the course administrator (content expert) and visible to all school staff (i.e. across schools) completing the SAS-WOC Online Teacher Training Course • Encourage teachers to offer peer support both during training and programme delivery.
Private Messaging	Private messaging function enabled within course	• Participants who had questions or queries they wished to ask the administrator/content expert privately or confidentially. • Messages were not visible to anyone other than the inquirer and the course administrator.

Functional testing of SAS-WOC online was conducted by novice users prior to operational release. The following testing protocols were employed:

i) **Mainline functions** – Functionality was tested by completing all modules, verifying the output and comparing the actual results with the expected results.
ii) **Basic Usability** – Was the user able to navigate through the training program without difficulty?
iii) **Accessibility** – Was the user able to access the system without difficulty?

The testing process verified that the SAS-WOC Online Teacher Training Course performed appropriately against functional requirements.

2.3.1.2 Completing SAS-WOC Online

An administrator-generated user name and system-generated password was forwarded via email to participants. Participants could choose to work through the program individually or in a group setting (sharing one device), however regardless of the preferred approach to accessing online content, a collegiate approach to learning was strongly encouraged. Colleagues were asked to completed training at the same time and in close proximity in order to allow training to be paused and discussion of key points and strategies to be held collectively.

2.3.1.3 Implementation and Evaluation

While it anticipated that future SAS-WOC online training would be available in a fully flexible format, in order to facilitate evaluation training was scheduled over two consecutive days with guidance placed in the course introduction (Module 1) as to the anticipated content goals for each day. A flexible approach to commencement times was adopted, however schools were advised of the times that the course administrator would be available online each day (8am – 4pm) and asked to complete their training within this window.

Following completion of the two-day training, participants (22 teachers and 7 teacher aides) were asked to provide an evaluation of the online training program. Participants rated the online training as ‘satisfactory’ with a mean of 5 on a scale of 1 to 7 (range 4-7). This is lower than evaluations of the face-to-face trainings. Notably, several participants did give very high ratings for the online training (7/7), with these teachers also being those that indicated that an online environment was their preferred training medium. Flexibility of access, pace and location were indicated as being the main reasons for this preference. Teachers who preferred an ‘online training’ environment found the video clips helpful and enjoyable and accessed the online administrator when they needed to do so. These comments are consistent with those made by teachers in other studies when asked about the benefits of online PD (Collins & Liang, 2015; Zhang, 2015).

3. Methodology

3.1 Design

The study was a cluster randomised controlled trial, with 613 students from 15 schools randomised to either wait-list or the Secret Agent Society—Whole of Class intervention (SAS-WOC). Randomisation was at the school level. Data was collected both from teachers and children at pre, post and 3-month follow-up (intervention only).

3.2 Participants

Table 8 shows the number of children with teacher-reported data available per school and time-point. Schools will be deidentified for reporting in publications. Eight schools were randomised to the wait-list control condition (*N* = 293 students) and seven schools were randomised to treatment (*N* = 320 students).

Demographic characteristics reported by teachers at baseline are summarised in Table 9. Overall, approximately half of the children participating were male (51.4%), with an average age of 10.5 years (SD = 0.52). A Welch's *t*-test indicated the average student age in the treatment group (*M* = 10.6, SD = 0.47) was significantly higher than the control group (*M* = 10.4, SD = 0.56), $t(569.13) = -3.89$, $p < .001$, however this finding is unlikely to be of clinical or practical importance.

The majority of students were reported by teachers at T1 as 'typically developing' (81.6%), with 11.1% reported as having socio-emotional difficulties and 7.3% reported to be on the autism spectrum. Few children (< 5% for each) were reported as having another diagnosis such as Anxiety, ADHD, learning disorder or a speech and language disorder.

Regarding child socio-emotional and behavioural indicators, almost one third of children were reported as having peer relationship difficulties at baseline (29.7%), and more than one-third reported to have difficulties staying on task in class (38.3%). Around one quarter were reported as being disruptive in class (25.4%). Teachers indicated that almost one in five students was bullied by other children (18.0%). Overall, the average number of classroom friends reported by teachers was 3.49 (SD = 2.28). The mean number of classroom friends was significantly higher for treatment schools (*M* = 3.76, SD = 2.47) relative to control schools (*M* = 3.21, SD = 2.02), $t(587.04) = -2.98$, $p = .003$, though again this statistical difference is unlikely to have practical significance.

Overall, around half the children (54.7%) were reported as being in the average range for academic achievement, while about one quarter were below average (24.5%). The chi-square test

of independence for *academic achievement* by *group* was significant, $\chi^2_{(2, N = 609)} = 14.69$, $p = < .001$. Treatment schools had fewer students proportionally in the below average range, and more in the average and above average ranges, than the wait-list schools. About one in five students (19.9%) was reported to be receiving special help in the classroom.

3.2 Measures

Data was collected from the following outcome measures at pre- and post-intervention:

3.2.1 Teacher Report

3.2.1.1 Social Skills Questionnaire (Teacher version) (Spence,1995; SSQ)

The Social Skills Questionnaire (SSQ-T) is a 30-item measure of children's social skills completed by teachers. Items are scored on a scale of 0 – 2 (0 = Not true; 1 = Sometimes true; 2 = Mostly true). All 30 items are summed to produce a SSQ Total score, with higher scores indicating more social skills. The measure showed strong internal consistency in the sample (α = .97).

3.2.1.2 Emotion-Regulation Social Skills Questionnaire (Teacher version) (Beaumont & Sofronoff, 2008; ERSSQ).

The Emotion-Regulation Social Skills Questionnaire Teacher version (ERSSQ-T) contains 25 items measuring students' emotion regulation and social skills and behaviours. Items are scored on a scale of 0 (Never) to 4 (Always). Two items (5 and 13) are reverse-scored prior to computing an ERSSQ Total score, so that higher Total scores indicate better social skills. Internal consistency in the current sample was good (α = .97).

Table 8: Number of children per school

Number of children per school based on teacher report.

School	Group	Pre (N)	Post (N)	All (N)
Bethany Lutheran Primary School	WL	29	25	54
Congupna Primary School	Trt	3	3	6
Darling Heights State School	Trt	27	26	53
Digger's Rest	WL	7	7	14
Eleebana	Trt	30	30	60
Holy Trinity	WL	27	27	54

School	Group	Pre (N)	Post (N)	All (N)
Livingstone CC	Trt	84	82	166
Murchison	Trt	14	14	28
Our Lady of the Southern Cross	WL	41	38	79
Pacific Paradise	WL	52	47	99
Shepparton East	WL	27	27	54
St Mary's Primary School	WL	56	55	111
St. Augustine's	Trt	88	77	165
Urangan Point	WL	54	51	105
Varsity College	Trt	74	71	145
(all)		613	580	1193

Note. WL = Wait-list; Trt = Treatment.

Table 9. Teacher-Reported Descriptive Statistics at Baseline

Description of students completing SAS-WOC Online

	All N=613	WL N=293	Trt N=320	*p*	Avail. *N*
Student gender: Male	315 (51.4%)	142 (48.5%)	173 (54.1%)	0.192	613
Student age (M±SD)	10.5±0.52	10.4±0.56	10.6±0.47	<0.001	612
Main diagnosis:				0.083	613
TD	500 (81.6%)	237 (80.9%)	263 (82.2%)		
ASD	45 (7.3%)	28 (9.6%)	17 (5.3%)		
SED	68 (11.1%)	28 (9.6%)	40 (12.5%)		
Other diagnosis: Anxiety	23 (3.8%)	12 (4.1%)	11 (3.4%)	0.829	613
Other diagnosis: ADHD	14 (2.3%)	9 (3.1%)	5 (1.6%)	0.328	613
Other diagnosis: Intellectual impairment	2 (0.3%)	1 (0.3%)	1 (0.3%)	1.000	613
Other diagnosis: Speech/language disorder	13 (2.1%)	6 (2.1%)	7 (2.2%)	1.000	613
Other diagnosis: Physical disorder	7 (1.1%)	4 (1.4%)	3 (0.9%)	0.715	613
Other diagnosis: Auditory processing	1 (0.2%)	1 (0.3%)	0 (0.0%)	0.478	613
Other diagnosis: Learning disorder	6 (1.0%)	4 (1.3%)	2 (0.6%)	0.433	613
Other diagnosis: Other	20 (3.3%)	10 (3.4%)	10 (3.1%)	1.000	613
No. of friends in class (M±SD)	3.49±2.28	3.21±2.02	3.76±2.47	0.003	601
Has peer relationship difficulties	182 (29.7%)	96 (32.8%)	86 (27.0%)	0.139	612
Academic achievement:				0.001	609

	All N=613	WL N=293	Trt N=320	*p*	Avail. *N*
Below average	149 (24.5%)	91 (31.4%)	58 (18.2%)		
Average	333 (54.7%)	147 (50.7%)	186 (58.3%)		
Above average	127 (20.9%)	52 (17.9%)	75 (23.5%)		
Has difficulty staying on task in class	235 (38.3%)	109 (37.2%)	126 (39.4%)	0.639	613
Is disruptive in class	156 (25.4%)	71 (24.2%)	85 (26.6%)	0.569	613
Child is bullied by others	109 (17.9%)	60 (20.7%)	49 (15.4%)	0.112	608
Receives special help in class	122 (19.9%)	64 (21.8%)	58 (18.1%)	0.294	613

Note. Avail. N refers to the number of non-missing responses in the data per variable. Proportions are computed based on non-missing responses.

3.2.1.3 Behaviour Assessment System for Children, Third Edition (Teacher version) Kamphaus & Reynolds, 2015)

The Teacher Rating Scale of the Behaviour Assessment System for Children, Third Edition (BASC-3) was used to assess students' behavioural and emotional difficulties. Items are rated *Never*, *Sometimes*, *Often*, *or Almost always*, which are converted to scores of 0 – 3.

The BASC-3 provides a comprehensive scoring procedure including community and clinical norms derived from *T*-scores. Raw scores were used for the current analyses to retain the focus on change over time rather than comparison to clinical norms. Items load onto 15 individual subscales assessing a broad range of domains (*Hyperactivity, Aggression, Conduct Problems, Anxiety, Depression, Somatization, Attention Problems, Learning Problems, Atypicality, Withdrawal, Adaptability, Social Skills, Leadership, Study Skills,* and *Functional Communication*). Subscales can be combined into five composite scales: (1) Externalising Problems (Hyperactivity, Aggression, and Conduct Problems); (2) Internalizing Problems (Anxiety, Depression, and Somatization); (3) School Problems (Attention Problems and Learning Problems); (4) Behavioural Symptoms (Atypicality and Withdrawal); and (5) Adaptive Skills (Adaptability, Social Skills, Leadership, and Functional Communication).

For the present study, the BASC-3 was not administered to students reported by teachers as 'typically developing'. Internal consistency (Cronbach's alpha) was computed for individual subscales and ranged between .80 and .93.

3.2.1.4 Anxiety Scale for Children—Autism Spectrum Disorder (ASC-ASD; Rodgers, Wigham, McConachie, Freeston, Honey & Parr, 2016).

The Anxiety Scale for Children—Autism Spectrum Disorder (ASC-ASD) contains 24 items assessing children's anxiety. This measure was administered to the subsamples of children reported to have socio-emotional difficulties or an Autism Spectrum Disorder.

Items on the ASC-ASD are rated on a scale of 0-3, with higher scores indicative of more anxiety. The measure produces a total score by summing all items together (α = .96), as well as four subscales: (1) Separation Anxiety (SA; α = .90); (2) Uncertainty (U; α = .92); (3) Performance Anxiety (PA; α = .84); and Anxious Arousal (AA; α = .90). As indicated, internal consistency in the sample was strong for all subscales and the Total score.

3.2.1.5 Child Adjustment and Parenting Efficacy Scale – Developmental Disability (Teacher version) (Mazzucchelli, Sanders, & Morawska, 2011)

The Child Adjustment and Parenting Efficacy Scale – Developmental Disability was adapted from the standard Child Adjustment and Parenting Efficacy Scale specifically for children with disability. The teacher version of the scale was administered in this study with regards to only those students reported to have socio-emotional difficulties or an Autism Spectrum Disorder.

The measure contains 13 items assessing intensity of child behavioural and emotional difficulties (CAPES-DD Intensity), with 3 of these items loading onto an emotional difficulties subscale and 10 loading onto a behavioural difficulties subscale. A further 8 items assess children's prosocial behaviours. Each item is scored on a scale of 0 (*Not at all*) to 3 (*Very much*), with the 13 Intensity items phrased such that higher scores indicate more difficulties. For the eight prosocial items, higher scores are positive, indicating more prosocial behaviours. Internal consistency for the Emotional, Behavioural, Intensity composite (Emotional plus Behavioural items), and Prosocial subscales were all good (α = .71, .88, .86, and .83 respectively).

In addition, the 13 Intensity items are rated by the respondent on a scale of 1 to 10 indicating their level of confidence in managing the difficult behaviour. These items are summed to produce an overall Self-Efficacy score, ranging from 13 to 130. Higher scores indicate more confidence. Internal consistency was also high for this subscale (α = .94).

3.3.3 Child Report

3.3.3.1 James and Dylan (Attwood, 2004)

Two written vignettes were presented to children describing a fictional character encountering a common anxiety-provoking scenario, such as a maths test (James) or being teased (Dylan). Children were given one open-ended question for each scenario asking them to describe how the

character should respond. Qualitative responses were reviewed by the lead researcher and given a score ranging between 0 and 9, reflecting how many appropriate emotion management strategies a student suggested for a vignette.

3.3.3.2 Bullying Prevalence Questionnaire

The Bullying Prevalence Questionnaire (BPQ) is a 20-item child self-report measure with three subscales: Bully, Victim and Prosocial. Items are rated from 1 to 4, with higher scores indicating more of the respective behaviour or difficulty. For the current sample, internal consistency was strong for the Bully (α = .80) and Victim subscales (α = .87), and adequate for the Prosocial subscale (α = .69).

4. Results

Data was analysed using *R*. Participants with T2 or T3 data but no T1 data were removed prior to analyses. For example, students that may have joined the classroom after the study had started and completed post-assessment but were not present in the original sample.

4.1 Time 1 – Time 2

A series of two-way, mixed ANOVAs were conducted to examine Group, Time and Group x Time interaction effects for Time 1 to Time 2. Missing data was largely due to attrition at T2. Based on the total number of T1 outcome responses available across all measures, 11.9% of responses were missing at T2 in total. However, the response rate varied across outcome measures even at T1. This variability may be due to random non-response, or the respondent omitting some subscale items which precluded calculation of subscale composite scores. For intent-to-treat analyses, missing subscale scores at T1 and T2 were imputed using EM imputation.

Analyses were conducted for teacher-reported ERSSQ and SSQ outcomes using the full available sample of students (n = 613). Analyses were re-run for these same measures, along with the CAPES-DD, BASC-3, and ASC outcomes, for the subsample of students designated as having ASD/SED (n = 113).

Group x Time interaction effects were reported for **all students (Child Report)** on the following outcomes:

- James & the Maths Test ($p < .001$)
- Dylan is being teased ($p = .003$)

Group x Time interaction effects were reported for **ASE-SED students only (Teacher Report)** on

- ASC U ($p = .043$)
- ASC Total ($p = .042$)
- BASC-3 Leadership ($p = .023$)

These results demonstrate a significant improvement in students in the intervention groups compared with students in the wait-list group. Table 10 shows means and standard deviations at T1 and T2 for each outcome measure, along with internal consistency (Cronbach's alpha). These were computed from the available data at each time point.

Table 10: Summary Statistics for Outcome Measures at T1 and T2

Means and standard deviations for outcome measures with internal consistency (Cronbach's Alpha)

		WL				Trt			
		Pre		Post		Pre		Post	
Outcome	α	*M*	SD	*M*	SD	*M*	SD	*M*	SD
Teacher data: All students (n = 613)									
SSQ Total (Social Skills Q)	0.97	50.64	11.63	51.06	11.59	50.31	11.59	51.58	10.75
ERSSQ Total (Emotion Reg & Soc Skills Q)	0.97	76.55	18.96	78.05	18.88	75.74	19.10	77.72	18.15
Teacher data: ASD/SED students only (n = 113)									
ASC PA	0.84	5.73	3.43	4.99	2.90	5.82	3.32	5.55	2.40
ASC AA	0.90	3.71	3.83	3.57	3.59	3.18	3.67	2.93	2.93
ASC SA	0.90	3.79	3.47	3.47	3.94	3.12	3.51	2.54	3.15
ASC U	0.92	7.36	5.43	6.77	5.39	5.98	5.16	6.12	4.62
ASC Total	0.96	20.43	15.07	18.70	15.03	18.02	13.52	16.67	9.27
CAPES-DD Emotional	0.71	2.50	2.03	2.35	2.08	2.87	2.08	2.17	1.62
CAPES-DD Behavioural	0.88	6.62	5.60	6.71	6.31	7.52	6.24	5.63	5.45
CAPES-DD Intensity	0.86	9.12	6.49	9.06	7.52	10.40	6.89	7.80	5.74
CAPES-DD Prosocial	0.83	13.00	4.55	13.17	4.59	13.35	4.28	13.33	4.15
CAPES-DD Self-Efficacy	0.94	117.23	16.27	121.38	12.32	116.82	17.17	117.73	17.41
Teacher data: BASC (n = 109)									
BASC-3 Adaptability	0.87	13.02	4.58	12.36	4.53	11.73	5.09	11.75	4.82
BASC-3 Aggression	0.93	5.88	5.70	6.34	5.60	7.11	6.96	6.65	6.68
BASC-3 Anxiety	0.90	9.22	5.27	8.70	4.91	10.93	6.12	10.47	5.65
BASC-3 Attention problems	0.93	11.02	5.78	10.57	5.74	14.64	5.02	13.22	5.06
BASC-3 Atypicality	0.82	7.35	4.48	7.02	4.00	8.11	4.27	7.45	4.57
BASC-3 Conduct	0.90	5.19	4.39	6.28	5.22	5.77	5.24	5.92	4.86
BASC-3 Depression	0.88	7.81	5.11	8.35	5.65	9.91	6.47	9.18	5.71
BASC-3 Functional communication	0.80	15.80	5.08	15.77	5.05	13.82	4.79	15.00	4.66
BASC-3 Hyperactivity	0.93	9.10	6.30	9.59	6.73	11.14	8.82	10.68	8.30
BASC-3 Leadership	0.80	6.80	4.02	6.15	3.51	4.68	2.87	5.55	3.53
BASC-3 Learning problems	0.93	9.33	6.14	9.06	5.99	12.64	6.87	11.28	6.50
BASC-3 Social skills	0.92	11.44	6.45	10.41	4.48	10.00	6.07	10.75	6.03
BASC-3 Somatization	0.90	3.84	4.03	4.13	5.26	3.80	4.66	4.47	5.05
BASC-3 Study skills	0.90	11.94	6.05	10.65	5.70	9.39	4.66	9.03	5.45
BASC-3 Withdrawal	0.85	8.71	4.42	9.09	3.56	9.95	4.33	9.47	3.49
BASC-3 Externalising composite		20.02	15.31	22.35	16.46	24.02	19.51	23.25	18.49
BASC-3 Internalising composite		20.90	12.63	21.11	13.68	24.64	13.79	24.12	12.87
BASC-3 School problems composite		20.35	11.02	19.64	10.87	27.27	10.59	24.50	10.42
BASC-3 Behavioural symptoms composite		16.08	8.03	16.04	6.78	18.07	7.18	16.93	6.76
BASC-3 Adaptive skills composite		58.55	22.79	55.41	19.65	49.61	18.92	52.08	20.52
Child self-report data (n = 570)									
BPQ: Bully	0.80	7.03	2.10	6.75	1.87	6.90	1.91	6.79	1.63
BPQ: Victim	0.87	9.29	3.40	8.87	3.41	8.52	3.34	8.46	3.09
BPQ: Prosocial	0.69	12.82	2.31	12.68	2.44	12.87	2.38	12.51	2.32
Dylan Q		1.57	1.01	1.88	1.15	1.63	0.94	2.28	1.25
James Q		1.08	0.84	1.44	1.07	1.29	0.87	2.16	1.21

Note. Cronbach's alpha not computed for BASC composite scales (as these were calculated by summing relevant subscales) and single-item measures (i.e. James and Dylan questions). WL = Wait-list group; Trt = Intervention group.

4.2 Time 1 – Time 3

Outcomes at T3 (three-month follow-up) relative to baseline for the *Secret Agent Society—Whole of Class* (SAS-WOC) were undertaken for the intervention cohort of the study. Analyses were conducted for teacher-reported ERSSQ and SSQ outcomes using the full available sample of intervention students (n = 287). Analyses were re-run for these same measures, along with the CAPES-DD, BASC-3, and ASC outcomes, for the subsample of intervention students designated as having ASD/SED (n = 50).

For the 287 students included in the follow-up analyses, T2 data was missing for around 7% of students for ERSSQ and 6% for SSQ at T2. At T3, 29% of data was missing for these subscales. For the ASD/SED sample, substantially higher proportions of responses were missing at T3. For the intent-to-treat analyses, missing T2 and T3 responses were imputed using EM imputation.

Follow-up analyses were conducted using repeated-measures ANOVA, with *time* entered as a within-subjects categorical variable, dummy-coded with two levels (representing T2 and T3) and with T1 as the baseline/reference category. Where necessary, the Greenhouse-Geiser correction was applied to address sphericity violations (i.e. non-homogenous variances across all combinations of T1-T3).

Where omnibus repeated-measures ANOVAs were significant, post hoc contrasts were conducted to examine pre-post and pre-follow-up differences. For post hoc comparisons, *p* values were adjusted using the Holm (1979) approach, which is a less conservative Type I error correction than the Bonferroni method.

Using the **whole sample** there are no further gains reported by teacher at T3. At T3, however, teachers reported significantly lower scores on the two measures and social skills and emotional regulation (SSQ and ERSSQ) compared with T1. This is a very unusual finding and is likely related to methodological issues and reporting issues. There was a high level of teacher turnover that has possibly affected the reporting of scores for typically developing students. This will be investigated further as part of the preparation of a paper.

In the **ASD/SED sample**, omnibus tests were significant for the ERSSQ (p =0.03) and SSQ (p=.041) outcomes, with mean scores significantly higher at both T2 and T3 relative to baseline. Follow-up tests following the significant omnibus test for CAPES-DD Emotional problems revealed that T2 scores were significantly lower than T1, however by T3 there was no significant difference from baseline scores. There was a significant ANOVA for the CAPES-DD Prosocial and BASC

Adaptability subscales, showing reductions from baseline levels at T3 but these effects were not evidenced at T2.

5. Discussion

The purpose of the SAS-WOC trial was to evaluate the feasibility and effectiveness of delivering a social-emotional skills program in an inclusive school setting to a whole classroom. To this end, the program was delivered by teachers following training and was modified to fit within a single term, with 45-minute weekly lessons. It was, therefore, both a light touch version of the SAS program and a universal delivery. Notably, earlier trials of the SAS program have been targeted interventions delivered in a small group setting with greater intensity (Beaumont & Sofronoff, 2008; Einfeld et al., 2018).

There is literature to suggest that universal programs when subjected to rigorous, standardised evaluation in a randomized controlled trial frequently fail to show compelling quantitative results (e.g., Sheffield et al., 2006; Spence et al., 2005). Similar to SAS-WOC, the large-scale trial undertaken by Sheffield et al. (2006) and Spence et al. (2005) also trained teachers to deliver a whole of classroom program but with the addition of evaluations extended over several years. The findings suggested minimal changes over time.

Some of the reasons for these findings are self-evident, the majority of students in the sample do not have social or emotional problems that are captured on the standardized measures (largely developed for a targeted population) and so there is little or no room for improvement. Universal programs are also usually lighter touch, less intense and brief, so those students who do have problems might not receive a large enough dose of the program to show an effect that can be captured by the measures. Having said that there is still a strong appetite in the education literature to deliver universal social-emotional programs. Several reviews have been completed of a diverse range of whole of classroom programs with mixed results (O'Conner, Dyson, Cowdell, & Watson, 2018; Sklad, Diekstra, de Ritter, Ben, & Gravestein, 2012). Author comments indicate a range of potential reasons including small and/or underpowered samples, high attrition, and wash out of effects at follow-up.

In the current trial there were few significant effects found for the whole sample of students. There was no significant effect found for social skills (SSQ) or for emotion regulation and social skills (ERSSQ). On the child questionnaires there was a significant effect shown for both the JAMES and DYLAN scenarios, which demonstrates that students did learn the strategies that could be used in situations evoking anxiety or anger.

Looking more closely at those students with a diagnosis on the autism spectrum or with teacher identified social-emotional difficulties, some significant changes are reported. For these students changes from pre-intervention to post-intervention were reported on the anxiety scale such that those students in the intervention group were seen as showing fewer anxious symptoms following the intervention compared with the students in the wait-list condition.

Furthermore, these students do show a significant improvement on teacher reported social skills (SSQ) and on teacher reported emotion regulation skills (ERSSQ). Changes from baseline to post and follow-up suggest that teachers are seeing improvement in these students. It is important to note however, that this analysis, was completed using all students after completion of the program - so a controlled comparison is not available. The finding for the larger group suggests that the trial was likely underpowered to detect these changes in the controlled condition.

5.1 Limitations

As mentioned above, it is likely that the sample size is small for the type of analyses required to answer the questions posed by the study. Given the nature of universal programs, a small effect would be anticipated and hence a very large sample is needed.

There are also some data considerations to be aware of in this study. One complication is that student IDs were manually entered by respondents, based on a unique combination of school, student initials, gender, and teacher or classroom details. In some cases, these IDs were entered inconsistently by school staff across time points, making it difficult to match data from the same student at both T1 and T2. This was compounded by the multi-informant design of the study (teacher and student report), resulting in separate datasets (teacher-report data and child-report data) that did not always align. In principle these assess the same set of students, however in practice this caused some difficulties when trying to match up teacher-report and child-report data.

The following approach was used to manage this:

- Student-report and teacher-report data were stored and analysed as separate datasets.
- Where IDs at T1 did not have a match at T2 (or vice versa), these were examined manually and corrected where possible, if reason for mismatch was clear. For example, some teachers entered their school initials as VC at T1 and VL at T2, resulting in mismatch. These cases were easily fixed.
- In other cases, the mismatch was more ambiguous. For example, if a teacher entered the student's initials as just JM at T2, while there is both JBM and JCM at T1. For schools with only a few students, these could sometimes be corrected without difficulty. However, if a

correction could not be made confidently, the student ID was left unchanged. Thus, there is the possibility that some matching pairs were missed and treated as missing data.

5.1.1 Student difficulties

'Student difficulties' was a teacher-reported variable used to categorise students as either "typically developing" (TD), having socio-emotional difficulties (SED), or having an Autism Spectrum Disorder (ASD). Teachers responded to this item at both T1 and T2, but in some cases provided a different value at each time point (e.g. TD at T1, SED at T2). As *student difficulties* was to be used as a moderator variable and to identify students for subgroup analyses, a single categorisation was needed for each student. For this study, we used the baseline value. So, students reported as having ASD or SED at T1 were designated as being in the ASD/SED subgroup, even if reported at T2 as being TD.

5.1.2 Staff changes

In some cases the teacher who completed the time 1 data and delivered the program and completed time 2 data was not available to complete follow-up. There were a considerable number of staff changes in the course of the program implementation resulting in loss of data. There were cases where a teacher left the school (through illness for example) and the program was then delivered by another teacher. In some instances, the person who delivered the program was not the classroom teacher. This would have a significant impact on the extent to which the teacher knew the students. It is likely that those students identified as on the autism spectrum or with social-emotional difficulties are known better and followed more closely by staff. This might be another reason for more changes being noticed in these students.

5.2 Conclusions

On the basis of the outcomes reported by teachers for those students on the autism spectrum and for those with social-emotional difficulties, it might be suggested that the program had some success. For this reason, and for individual student responses not captured by the standardized measures, several schools have continued to implement the SAS-WOC program in the school. When we look at the satisfaction ratings provided by students after completing the program, a significant percentage (72%) reported finding the SAS-WOC program both enjoyable and helpful with a majority of students (87%) able to identify specific strategies that they found helpful and used. The question that remains is whether the program as it is currently configured demonstrates sufficient effectiveness to suggest it be disseminated.

5.2.1 Future directions

If teaching social-emotional skills is to become a real focus in mainstream schools, as suggested in many papers (e.g. Humphrey et al., 2016) and book chapters (e.g., Daly, Nicholls, Aggarwal & Sander, 2014), then it must also be given 'real time' not just in the curriculum but in the school timetable itself. The stated objectives of most social-emotional programs do align with National Curriculums in many countries. However, in many schools this is not seen as 'core business' and is therefore managed in a piecemeal way rather than with a dedicated program of learning. In many other schools, where social-emotional skill development is seen as core business, there is often no free time that can be scheduled in the timetable.

In terms of concerns specific to the SAS-WOC program, the issue of dose was raised by teachers in their feedback to us. Many teachers would have preferred to deliver the program over a whole year rather than the one term required to fit within the research protocol. This could easily be done and would allow skills and strategies to be introduced more gradually and practice to be carried out over a longer time. Teachers also found that their students were very keen to discuss issues raised by the program such as those of individual differences, helping others, diversity of all types and so on. The discussion often meant that the classes ran over time or (alternatively) discussion had to be cut short resulting in lost learning opportunities. With the program more evenly spread out this could be accommodated.

The way that the computer game was used was also found to be important. Students reported greater involvement and enjoyment of the program when they were able to access the computer game individually and work at their own pace. They reported enjoying the program least when the game was controlled by the teacher. Since the game can be streamed by schools, there is no reason for students to not have individual access. This would allow some components to be completed individually and other components to be completed in groups within the classroom.

In a majority of schools that participated in the trial, there was at least one staff member who had been trained in the SAS small group program. This was frequently a school counsellor or guidance officer. None of the students in the year 5 classrooms participating in the trial had completed the small group program. Given the need for greater rehearsal of skills and poor generalization of skills in children on the autism spectrum, it would potentially be beneficial to offer the small group program to those students with greater need and to reinforce the skills and strategies in a much lighter touch program suitable for all students. The utility of this approach could be the focus of a future trial.

6. References

Achenbach, T. M. (1991). Manual for the Child Behavior Checklist/4-18 and 1991 Profile University of Vermont. *Department of Psychiatry, Burlington*.

Achenbach, T. M., Dumenci, L., & Rescorla, L. A. (2001). Ratings of relations between DSM-IV diagnostic categories and items of the CBCL/6-18, TRF, and YSR. *Burlington, VT: University of Vermont*.

Andrews, L., Attwood, T., & Sofronoff, K. (2013). Increasing the appropriate demonstration of affectionate behavior, in children with Asperger syndrome, high functioning autism, and PDD-NOS: A randomized controlled trial. *Research in autism spectrum disorders*, *7*(12), 1568-1578.

Asher, S. R., Hymel, S., & Renshaw, P. D. (1984). Loneliness in children. *Child development*, 1456-1464.

Attwood, T. (2004). Cognitive Behaviour Therapy for Children and Adults with Asperger's Syndrome. *Behaviour Change,21*(3), 147-161.

Attwood, T. (2004a). James and the maths test. In *Exploring feelings: Cognitive behaviour therapy to manage anger*, 65-66. Arlington, TX: Future Horizons.

Attwood, T. (2004b). Dylan is being teased. In *Exploring feelings: Cognitive behaviour therapy to manage anger*, 65-66. Arlington, TX: Future Horizons.

American Psychiatric Association. DSM-5 Task Force, & American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders: DSM-5 (5th ed.). Washington, DC: American Psychiatric Association.

Baron-Cohen, S., Golan, O., Wheelwright, S., & Hill, J. J. Mind Reading: The Interactive Guide to Emotions. 2004. *London: Jessica Kingsley Ltd Google Scholar*.

Bates, M.S., Phalen, L., & Moran, C. (2016). Online professional development: A primer. Phi Delta Kappan, 97(5), 70-73. DOI: 10.1177/0031721716629662

Beaumont, R.B., Pearson, R., & Sofronoff, K. (submitted Aug. 2018). A novel intervention for peer relationship difficulties in children: The Secret Agent Society program. Journal of Child and Family Studies.

Beaumont, R., Rotolone, C., & Sofronoff, K. (2015). The secret agent society social skills program for children with high-functioning autism spectrum disorders: A comparison of two school variants. Psychology in the Schools, 52(4), 390-402. doi:10.1002/pits.21831

Beaumont, R., & Sofronoff, K. (2008). A multi-component social skills intervention for children with asperger syndrome: The junior detective training program. *Journal of Child Psychology and Psychiatry, 49(7),* 743-753. doi:10.1111/j.1469-7610.2008.01920.x

Beaumont, R., & Sofronoff, K. (2013). Multi-modal intervention for social skills training in students with Asperger syndrome: The Secret Agent Society. In A. Scarpa, S. White and T. Attwood (Eds), Promising cognitive behavioural interventions for children and adolescents with high functioning autism spectrum disorders. Guilford Publications.

Beaumont, R.B., Walker, H., Weiss, J., & Sofronoff, K. (submitted Aug. 2018). Improving access to a social skills program for children with autism spectrum conditions: A randomized controlled trial of a parent-directed version of the Secret Agent Society social-emotional skills program. Journal of Child Psychology and Psychiatry.

Bird, H. R., Shaffer, D., Fisher, P., & Gould, M. S. (1993). The Columbia Impairment Scale (CIS): Pilot findings on a measure of global impairment for children and adolescents. *International Journal of Methods in Psychiatric Research*.

Bukowski, W. M., Hoza, B., & Boivin, M. (1994). Measuring friendship quality during pre-and early adolescence: The development and psychometric properties of the Friendship Qualities Scale. *Journal of social and Personal Relationships*, *11*(3), 471-484.

Cappadocia, M.C., & Weiss, J. A. (2011). Review of social skills training groups for youth with Asperger syndrome and high functioning autism. *Research in Autism Spectrum Disorders, 5,* 70-78.

Casagrande, K., & Ingersoll, B. (2017). Service Delivery Outcomes in ASD: Role of Parent Education, Empowerment, and Professional Partnerships. *Journal of Child And Family Studies,26*(9), 2386-2395.

Chambers, D., Rodgers, M., Golder, S., Pepper, C., Todd, D., & Woolacott, N. (2009). *Systematic reviews: CRD's guidance for undertaking reviews in health care*. Centre for Reviews and Dissemination.

Chorpita, B. F., Yim, L., Moffitt, C., Umemoto, L. A., & Francis, S. E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: A revised child anxiety and depression scale. *Behaviour research and therapy*, *38*(8), 835-855.

Colman, A. (2008). Demand characteristics. *A Dictionary of Psychology,* A Dictionary of Psychology.

Constantino, J. N., & Gruber, C. P. (2012). *Social responsiveness scale (SRS)*. Torrance, CA: Western Psychological Services.

Desimone, L.M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. Educational Researcher, 38(3), 181–199. DOI: 10.3102/0013189X08331140

Einfeld, S., Beaumont, R., Clark, T., Clarke, K., Costley, D., Gray, K.M., Horstead, S.K., Redoblado Hodge, M.A., Roberts, J., Sofronoff, K., Taffe, J., & Howlin, P. (2018). School-based social skills training for young people with autism spectrum disorders. Journal of Intellectual and Developmental Disabilities. DOI: 10.3109/13668250.2017.1326587

Fisher, J. B., Schumaker, J. B., Culbertson, J., & Deshler, D. D. (2010). Effects of a

computerized professional development program on teacher and student outcomes. Journal of Teacher Education, 61(4), 302-312. DOI: 10.1177/0022487110369556

Fishman, B. J., Konstantopoulos, S., Kubitskey, B. W., Vath, R., Park, G., Johnson, H., &

Edelson, D. C. (2013). Comparing the impact of online and face-to-face professional development in the context of curriculum implementation. Journal of Teacher Education, 64(5), 426-438. DOI: 10.1177/0022487113494413

Frankel, F., & Mintz, J. (2011). Maternal Reports of Play Dates of Clinic Referred and Community Children. *Journal of Child and Family Studies, 20*(5), 623-630.

Frankel, F., & Mintz, J. (2008). Measuring the quality of play dates. *Available from UCLA Parenting and Children's Friendship Program*, *300*.

Frankel, F. H., & Myatt, R. (2003). *Children's friendship training*. Psychology Press.

Frankel, F., Myatt, R., Sugar, C., Whitham, C., Gorospe, C. M., & Laugeson, E. (2010). A randomized controlled study of parent-assisted children's friendship training with children having autism spectrum disorders. *Journal of autism and developmental disorders*, *40*(7), 827-842.

Gates, J. A., Kang, E., & Lerner, M. D. (2017). Efficacy of group social skills interventions for youth with autism spectrum disorder: A systematic review and meta-analysis. Clinical Psychology Review.

Golan, O., Baron-Cohen, S., & Hill, J. (2006). The Cambridge mindreading (CAM) face-voice battery: Testing complex emotion recognition in adults with and without Asperger syndrome. *Journal of autism and developmental disorders*, *36*(2), 169-183.

Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating system: Manual*. American Guidance Service.

Guy, W. (1976). CGI clinical global impressions. *EC-DEU Assessment Manual for Psychopharmacology*, 76-338.

Hinton, S., Sofronoff, K., & Sheffield, J. (2008) Training teachers to manage students with Asperger's Syndrome in an inclusive classroom setting. Australian Educational and Developmental Psychologist, *25*(2) 34-48.

Howlin, P., & Taylor, J. (2015). Addressing the need for high quality research on autism in adulthood. Autism, 19(7), 771

IBM Corp. (2013). IBM SPSS Statistics for Macintosh, Version 22.0. Armonk, NY: IBM Corp.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of consulting and clinical psychology*, *59*(1), 12.

Koning, C., Magill-Evans, J., Volden, J., & Dick, B. (2013). Efficacy of cognitive behavior therapy-based social skills intervention for school-aged boys with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *7*(10), 1282-1290.

Koning, C., Magill-Evans, J., & Volden, J. (2008, May). Measurement of peer interaction in children with HFA. In *Poster presented at the International Meeting for Autism Research, London*.

Kovacs, M. (1985). The children's depression inventory (CDI). *Psychopharmacol bull*, *21*, 995-998.

La Greca, A. M., & Lopez, N. (1998). Social anxiety among adolescents: Linkages with peer relations and friendships. *Journal of abnormal child psychology*, *26*(2), 83-94.

Laugeson, E. A., Ellingsen, R., Sanderson, J., Tucci, L., & Bates, S. (2014). The ABC's of teaching social skills to adolescents with autism spectrum disorder in the classroom: The UCLA PEERS® program. *Journal of autism and developmental disorders*, *44*(9), 2244-2256.

Laugeson, E. A., & Frankel, F. (2010). Social skills for teenagers with developmental and autism spectrum disorders: The PEERS treatment manual. New York, NY: Routledge.

Laugeson, E. A. & Frankel, F. (Eds.). (2010). Test of Adolescent Social Skills Knowledge, Mental Status Checklist. In the PEERS treatment manual. New York: Routledge.

Laugeson, E., & Park, A. (2014). Using a CBT Approach to Teach Social Skills to Adolescents with Autism Spectrum Disorder and Other Social Challenges: The PEERS ® Method. Journal of Rational-Emotive & Cognitive-Behavior Therapy, 32(1), 84-97.

Laugeson, E. A., Frankel, F., Gantman, A., Dillon, A. R., & Mogil, C. (2012). Evidence-based social skills training for adolescents with autism spectrum disorders: The UCLA PEERS program. *Journal of autism and developmental disorders*, *42*(6), 1025-1036.

Lawrence D, Johnson S, Hafekost J, Boterhoven De Haan K, Sawyer M, Ainley J, Zubrick SR (2015) The Mental Health of Children and Adolescents. Report on the second Australian Child and Adolescent Survey of Mental Health and Wellbeing. Department of Health, Canberra.

Lopata, C., Thomeer, M. L., Volker, M. A., Toomey, J. A., Nida, R. E., Lee, G. K., ... & Rodgers, J. D. (2010). RCT of a manualized social treatment for high-functioning autism spectrum disorders. *Journal of autism and developmental disorders*, *40*(11), 1297-1310.

Lord, C., & Rutter, M. (2003). Social communication questionnaire (SCQ). *Torrance, CA: WPS*.

Lounds, J. T., Dove, D., Veenstra-VanderWeele, J., Sathe, N. A., McPheeters, M. L., Jerome, R. N., & Warren, Z. (2012). Interventions for adolescents and young adults with autism spectrum disorders.

Loveland, K. A. (2005). Social-emotional impairment and self-regulation in autism spectrum disorders. *Emotional development*, 365-382.

Lubans, D., Plotnikoff, R., & Lubans, J. (2012). A systematic review of the impact of physical activity programmes on social and emotional well-being in at-risk youth. *Child and Adolescent Mental Health*, *17*(1), 2-13.

Magill-Evans, J., Koning, C., Cameron-Sadava, A., & Manyk, K. (1995). The child and adolescent social perception measure. *Journal of Nonverbal Behavior*, *19*(3), 151-169.

March, J. S., Parker, J. D., Sullivan, K., Stallings, P., & Conners, C. K. (1997). The Multidimensional Anxiety Scale for Children (MASC): factor structure, reliability, and validity. *Journal of the American academy of child & adolescent psychiatry*, *36*(4), 554-565.

Masters, J., de Kramer, R. M., O'Dwyer, L. M., Dash, S., & Russell, M. (2012). The effects

of online teacher professional development on fourth grade students' knowledge and practices in English language arts. Journal of Technology and Teacher Education, 20(1), 21-46.

Mattick, R. P., & Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety1. *Behaviour research and therapy*, *36*(4), 455-470.

McGinnis, E., & Goldstein, A. P. (1997). Skillstreaming the elementary school child: New strategies and perspectives for teaching prosocial skills. Research Press.

Mendelson, J. L., Gates, J. A., & Lerner, M. D. (2016). Friendship in school-age boys with autism spectrum disorders: A meta-analytic summary and developmental, process-based model. *Psychological bulletin*, *142*(6), 601.

Montori, V. M., & Guyatt, G. H. (2001). Intention-to-treat principle. *Canadian Medical Association Journal*, *165*(10), 1339-1341.

Moree, B.N., & Davis, T.E. (2010). Cognitive-behavioral therapy for anxiety in children with autism spectrum disorders: Modification trends. Research in Autism Spectrum Disorders, 4, 346-354.

Nauta, M. H., Scholing, A., Rapee, R. M., Abbott, M., Spence, S. H., & Waters, A. (2004). A parent-report measure of children's anxiety: psychometric properties and comparison with child-report in a clinic and normal sample. *Behaviour research and therapy*, *42*(7), 813-839.

Nowicki Jr, S., & Duke, M. P. (1998). Manual for the receptive tests of the Diagnostic Analysis of Nonverbal Accuracy 2. *Atlanta, GA: Department of Psychology, Emory University*.

Pearson Clinical Assessment (2012). Behavior Assessment Q Global Build Version 54.2.11. *Pearson Education Inc.* accessed from https://qglobal.pearsonclinical.com/qg/au/login.seam

Pellecchia, M., Connell, J. E., Kerns, C. M., Xie, M., Marcus, S. C., & Mandell, D. S. (2015). Child characteristics associated with outcome for children with autism in a school-based behavioral intervention. *Autism : The International Journal of Research and Practice*, 1362361315577518-. http://doi.org/10.1177/1362361315577518

Picci, G., & Scherf, K. S. (2015). A two-hit model of autism: Adolescence as the second hit. *Clinical Psychological Science*, *3*(3), 349-371.

Pickard, K., & Ingersoll, B. (2016). Quality versus quantity: The role of socioeconomic status on parent-reported service knowledge, service use, unmet service needs, and barriers to service use. *Autism, 20*(1), 106-115.

Piers, E. V. (1984). Piers-Harris Children's Self-Concept Scale-revised manual Los Angeles: Counselor Psychological Tests. *WPS.[Links]*.

Prestridge, S., & Tondeur, J. (2015). Exploring elements that support teachers' engagement in online professional development. Education Sciences, 5(3), 199-219. doi:10.3390/educsci5030199

Reichow, B, & Volkmar, F. (2010). Social Skills Interventions for Individuals with Autism: Evaluation for Evidence-Based Practices within a Best Evidence Synthesis Framework. Journal of Autism and Developmental Disorders, 40(2), 149-166

Research Units on Pediatric Psychopharmacology Anxiety Study Group.(RUPP) (2002). The pediatric anxiety rating scale (PARS): Development and psychometric properties. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*(9), 1061-1069.

Reuters, T. (2013). EndNote X7. *Philadelphia: Thomson Reuters*.

Reynolds, C. R. (2004). *Behavior assessment system for children*. John Wiley & Sons, Inc..

Reynolds, C. & Kamphaus, R. (2002). *The clinician's guide to the Behavior Assessment System for Children (BASC)*. Guilford Press.

Reynolds, C. & Kamphaus, R. (2015). *BASC-3: Behavioral Assessment System for Children* (Third ed.). Bloomington, Minn.: PsychCorp.

Reynolds, C. R., & Richmond, B. O. (1985). *Revised children's manifest anxiety scale*. Los Angeles: Western Psychological Services.

Schohl, K. A., Van Hecke, A. V., Carson, A. M., Dolan, B., Karst, J., & Stevens, S. (2014). A replication and extension of the PEERS intervention: examining effects on social skills and social anxiety in adolescents with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *44*(3), 532-545.

Schroeder, H. (2016). Four essential tips for professional development success. eLearn Magazine. http://elearnmag.acm.org/featured.cfm?aid=3018842

Silverman, W. K., & Albano, A. M. (1996). The anxiety disorders interview schedule for children (ADIS-C/P). *San Antonio, TX: Psychological Corporation*.

Simonoff, E., Pickles, A., Charman, T., Chandler, S., Loucas, T., & Baird, G. (2008). Psychiatric Disorders in Children with Autism Spectrum Disorders: Prevalence, Comorbidity, and Associated Factors in a Population-Derived Sample. *Journal of the American Academy of Child & Adolescent Psychiatry, 47*(8), 921.

Sofronoff, K., Beaumont, R., & Weiss, J. (2015). Treating transdiagnostic processes in ASD: Going beyond anxiety. In Thompson E, Davis III, Susan W. White & Thomas H. Ollendick (Eds), *Handbook of autism and anxiety.* Springer Publishing Co. New York.

Sofronoff, K., Silva, J., & Beaumont, R. (2015). The Secret Agent Society social-emotional skills programme for children with high-functioning autism spectrum disorder: A parent-directed trial. Focus on Autism and Other Developmental Disorders, 1-16. DOI:10:1177/1088357615583467

Sparrow, S. S., & Cicchetti, D. V. (1985). Diagnostic uses of the vineland adaptive behavior scales. *Journal of Pediatric Psychology*, *10*(2), 215-225.

Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). Vineland adaptive behavior scales:(Vineland II), survey interview form/caregiver rating form. *Livonia, MN: Pearson Assessments*.

Spielberger, C. D., & Edwards, C. D. (1973). *State-trait anxiety inventory for children: STAIC: How I feel questionnaire: Professional manual*. Mind Garden.

Spence, S. H. (1995). Social skills training: Enhancing social competence with children and adolescents. Windsor, UK: Nfer-Nelson.

Storch, E. A., Arnold, E. B., Lewin, A. B., Nadeau, J. M., Jones, A. M., De Nadai, A. S., ... & Murphy, T. K. (2013). The effect of cognitive-behavioral therapy versus treatment as usual for anxiety in children with autism spectrum disorders: A randomized, controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, *52*(2), 132-142.

Storch, E. A., Lewin, A. B., Collier, A. B., Arnold, E., De Nadai, A. S., Dane, B. F., ... & Murphy, T. K. (2015). A randomized controlled trial of cognitive-behavioral therapy versus treatment as usual for adolescents with autism spectrum disorders and comorbid anxiety. *Depression and anxiety*, *32*(3), 174-181.

Sukhodolsky, D. G., Scahill, L., Gadow, K. D., Arnold, L. E., Aman, M. G., McDougle, C. J., ... & Vitiello, B. (2008). Parent-rated anxiety symptoms in children with pervasive developmental

disorders: Frequency and association with core autism symptoms and cognitive functioning. *Journal of abnormal child psychology*, *36*(1), 117-128.

Tabachnick, B., & Fidell, L. (2013). *Using multivariate statistics* (6th ed., New International ed.). Harlow: Pearson Education Limited.

Thomeer, M. L., Lopata, C., Volker, M. A., Toomey, J. A., Lee, G. K., Smerbeck, A. M., ... & Smith, R. A. (2012). Randomized clinical trial replication of a psychosocial treatment for children with high-functioning autism spectrum disorders. *Psychology in the Schools*, *49*(10), 942-954.

Thomeer, M. L., Rodgers, J. D., Lopata, C., McDonald, C. A., Volker, M. A., Toomey, J. A., ... & Gullo, G. (2011). Open-trial pilot of mind reading and in vivo rehearsal for children with HFASD. *Focus on Autism and Other Developmental Disabilities*, *26*(3), 153-161.

Thomeer, M. L., Smith, R. A., Lopata, C., Volker, M. A., Lipinski, A. M., Rodgers, J. D., ... & Lee, G. K. (2015). Randomized controlled trial of mind reading and in vivo rehearsal for high-functioning children with ASD. *Journal of autism and developmental disorders*, *45*(7), 2115-2127.

Wagner, A., Lecavalier, L., Arnold, L. E., Aman, M. G., Scahill, L., Stigler, K. A., ... & Vitiello, B. (2007). Developmental disabilities modification of the Children's Global Assessment Scale. *Biological Psychiatry*, *61*(4), 504-511.

Weiss, J. (2014). Transdiagnostic Case Conceptualization of Emotional Problems in Youth with ASD: An Emotion Regulation Approach. *Clinical Psychology: Science and Practice, 21*(4), 331-350.

Whear, R., Marlow, R., Boddy, K., Ukoumunne, O. C., Parker, C., Ford, T., ... & Stein, K. (2014). Psychiatric disorder or impairing psychology in children who have been excluded from school: a systematic review. *School Psychology International*, *35*(5), 530-543.

White, S., Albano, A. M., Johnson, C. R., Kasari, C., Ollendick, T., Klin, A., & Scahill, L. (2010). Development of a cognitive-behavioral intervention program to treat anxiety and social deficits in teens with high-functioning autism. Clinical Child and Family Psychology Review, 13, 77–90.

White, S., Keonig, K., & Scahill, L. (2007). Social skills development in children with autism spectrum disorders: A review of the intervention research. Journal of autism and developmental disorders, 37(10), 1858-1868.

White, S., Ollendick, T., Albano, A. M., Oswald, D., Johnson, C., Southam-Gerow, M. A., ... & Scahill, L. (2013). Randomized controlled trial: multimodal anxiety and social skill intervention for adolescents with autism spectrum disorder. Journal of autism and developmental disorders, 43(2), 382-394.

White, S. W., Oswald, D., Ollendick, T., & Scahill, L. (2009). Anxiety in children and adolescents with autism spectrum disorders. *Clinical psychology review*, *29*(3), 216-229.

Wise, E. A. (2004). Methods for analyzing psychotherapy outcomes: A review of clinical significance, reliable change, and recommendations for future directions. *Journal of personality assessment*, *82*(1), 50-59.

Wood, J. J., & Drahota, A. (2005). Behavioral interventions for anxiety in children with autism. *Los Angeles, CA: University of California–Los Angeles*.

Wood, J. J., Ehrenreich-May, J., Alessandri, M., Fujii, C., Renno, P., Laugeson, E., ... & Murphy, T. K. (2015). Cognitive behavioral therapy for early adolescents with autism spectrum disorders and clinical anxiety: A randomized, controlled trial. *Behavior Therapy*, *46*(1), 7-19.

Yoo, H. J., Bahn, G., Cho, I. H., Kim, E. K., Kim, J. H., Min, J. W., ... & Cho, S. (2014). A randomized controlled trial of the Korean version of the PEERS® parent-assisted social skills training program for teens with ASD. *Autism Research*, *7*(1), 145-161.

Zheng, L. (2015). In-service professional development in an online environment: What are South Australian English as an additional language or dialect teachers' view? Professional Development in Education, 41(3), 527-545. DOI: 10.1080/19415257.2014.902860

Autism CRC
The University of Queensland
Long Pocket Precinct
Level 3, Foxtail Building
80 Meiers Road
Indooroopilly QLD 4068

T +61 7 3377 0600
E info@autismcrc.com.au
W autismcrc.com.au

@autismcrc

Australian Government
Department of Industry, Innovation and Science

Business
Cooperative Research
Centres Programme