Understanding Social Behaviour in Children and Adolescents.

Joanna Loughrey

BSc (Hons) Psychology, MSc Applied Psychology

Submitted in partial fulfilment of the requirements for the degree of

Doctorate in Clinical Psychology

School of Health and Wellbeing

College of Medical, Veterinary and Life Sciences

University of Glasgow

July 2024

Table of Contents

List of Tables

List of Figures

Chapter 1: Systematic Review

Acknowledgements

Firstly, I would like to thank the parents and children who participated in the ABCD study and the ABCD research team. A sincere thank you to my research supervisors Professor Liam Dorris, Dr Breda Cullen and Dr Claire Adey for your guidance and support throughout all aspects of the project. I would also like to thank Dr Paul Cannon from the University of Glasgow library for your support in the development of my systematic review search strategy. Special thanks to my colleague and dear friend Rosa Sparks for being my second reviewer for my systematic review. Thank you to all my family and friends for all your love, encouragement and understanding throughout my training, particularly over these last few months. I can't wait to spend more time with you all. To my father Dominic, thank you for your regular phone calls, visits, words of encouragement and care packages, and all your support and love throughout my life. I am very lucky and grateful to have you as my father. A special thank you to my mother Catherine for your help and guidance with the countless pieces of coursework and interview prep that has gotten me to this stage, I would not have got here without you. Thank you also for instilling me with the self-belief to pursue my dreams. You are the most tenacious and resilient person I know, and it is a privilege to be your daughter. Lastly, a special thanks to my husband Peter for following me over to Glasgow to complete this doctorate in the midst of planning our wedding. Thank you for your endless support, patience and kindness. Thank you also for being my personal IT technician and always being on hand to help me with the endless IT and formatting issues I have encountered not only during this degree, but the other two. I promise, this will be my last degree.

Foreword

Due to not gaining educational approval, the original proposal for the Major Research Project did not go ahead (see Appendix M2.)

Chapter 1

A Systematic Review of the Relationship between Empathy and Cyberbystander Behaviour in Children and Adolescents

Prepared in accordance with the author requirements for British Journal of Psychology.
https://bpspsychub.onlinelibrary.wiley.com/hub/journal/20448295/homepage/forauthors.html

Abstract

Cyberbullying is a social and public health issue (Henares-Montiel et al., 2023). Research has begun to focus on cyberbystander behaviour and the factors that increase defending bystander behaviour and reduce passive and/or reinforcing bystander behaviour in a cyberbullying context. A positive significant relationship between empathy and defending bystander behaviours has been identified in previous systematic reviews on face-to-face bullying (Deng et al., 2021). A previous review on cyberbystander behaviour and empathy was inconclusive due to a lack of studies (Zych et al., 2019). Given the recent increase in interest in this topic, this review aimed to describe the relationship between empathy and cyberbystander behaviour in children and adolescents, and to examine if there is evidence to suggest cognitive *or* affective empathy has more influence on cyberbystander behaviour. Five databases were searched in addition to forward and backward chaining, final searches were completed on the 26th of June 2024. Eight articles were included and appraised for quality using the Crowe Critical Appraisal Tool. Seven of these articles were included in a narrative synthesis. Results from the included studies suggest a statistically significant but weak relationship between higher empathy and defending cyberbystander behaviour. Strengths, limitations and recommendations for future research are discussed.

References

Deng, X., Yang, J., & Wu, Y. (2021). Adolescent empathy influences bystander defending in school bullying: A three-level meta-analysis. Frontiers in psychology, 12, 690898

Henares-Montiel, J., Pastor-Moreno, G., Ramírez-Saiz, A., Rodriguez-Gomez, M., & Ruiz-Pérez, I. (2023). Characteristics and effectiveness of interventions to reduce cyberbullying: a systematic review. Frontiers in Public Health, 11, 1219727.

Zych, I., Baldry, A. C., Farrington, D. P., & Llorent, V. J. (2019). Are children involved in cyberbullying low on empathy? A systematic review and meta-analysis of research on empathy versus different cyberbullying roles. *Aggression and violent behavior*, *45*, 83-97.

Introduction

Cyberbullying is defined as ‘a set of behaviours performed through electronic or digital media by one individual or group of individuals who repeatedly communicate hostile or aggressive messages intended to inflict harm or discomfort on others’ (Tokunaga, 2010). A recent systematic review and meta-analysis measured the prevalence rates of cyberbullying in children and adolescents across 63 countries and established the prevalence of victimisation from cyberbullying ranges from 14% to 58%, and that the perpetration prevalence ranges from 6% to 46% (Zhu et al., 2021).

Cyberbullying develops differently from traditional bullying, with a typical onset around age 12-13 years (Li et al., 2022). Researchers have applied Bronfenbrenner’s (1979) ecological systems framework to understand cyberbullying perpetration and why it differs from traditional bullying. Traditional bullying tends to decline during middle and late adolescence whereas cyberbullying often increases during these years (Li et al., 2022). This trend can be attributed to changes in adolescents ‘meso-system’ whereby during middle adolescence parents and carers facilitate access to smartphones and social media, which provides their children with the opportunity to engage in cyberbullying (Hong et al., 2023). Additionally, because cyberbullying occurs online, parents have a more significant role compared to traditional bullying. Reduced parental supervision of internet use as children grow older likely contributes to the later onset of cyberbullying (Hong et al., 2023)

Another meta-analysis including 57 empirical studies from 17 countries reported a significant correlation between experiencing cyberbullying and depression, particularly in adolescents (Hu et al., 2021). Research has also shown being cyberbullied is associated with general psychological distress, academic difficulties, loneliness and reduced wellbeing (Aparisi et al., 2021; Jaing et al., 2022). Given the increased prevalence of this phenomenon amongst youths and the seriousness of its implications for health, cyberbullying has been identified as a social and public health issue (Henares-Montiel et al., 2023).

As a result of these detrimental effects, research has aimed to understand cyberbullying, particularly the behaviour of cyberbully perpetrators (Allison & Bussey, 2016) but has largely neglected the bystander's role in cyberbullying (Huang et al., 2019). A systematic review of cyberbystander behaviour in adolescents identified bystanders act in various ways, including active defending, privately supporting the victim, reinforcing the cyberbully or acting passively (Polanco-Levicán & Salvo-Garrido, 2021). Like 'traditional bullying' which is also described as 'face-to-face' or 'offline bullying' in the literature, research has reported defending behaviours which include actively intervening and telling the perpetrators to stop, can reduce the impact cyberbullying has on the victim (Allison & Bussey, 2016). The literature often refers to these types of behaviours as 'pro-social' cyberbystander behaviour.

Unfortunately, research suggests that the majority of cyberbystanders choose to remain passive when they witness cyberbullying. One study reported 69% of cyberbystanders took a passive role when they witnessed cyberbullying and did nothing (Song & Oh, 2018).

In addition to defending a cybervictim, or acting passively, research has found cyberbystanders may become a ‘cyberbully reinforcer’. Reinforcing behaviour includes liking a negative comment that a cyberbully has posted or re-sharing unkind content. Given that most adolescents will remain passive, and not engage in helping or defending behaviours, a research focus has been to understand the interpersonal and contextual factors that increase defending cyberbystander behaviour and reduce passive or reinforcing cyberbystander behaviour.

Researchers have suggested that to actively intervene, the bystander needs to feel a sense of connection with the victim and feel safe that they would not be victimised too (Obermaier et al., 2016). A systematic review and meta-analysis involving 25 studies confirmed a positive statistically significant relationship between affective and cognitive empathy and bystander defending behaviour in incidences of traditional bullying (Deng et al., 2021). Furthermore, the review reported affective empathy had a significantly stronger relationship with defending bystander behaviour compared to cognitive empathy in traditional bullying. If similar results were replicated in cyberbullying, this would inform the design of cyberbullying intervention and reduction programs. However, researchers have suggested it is more difficult to foster connection online, compared to ‘real life’ for various reasons, for example, seeing the expression of ‘hurt’ prompts defending behaviour in bystanders (Kozubal et al., 2019), which is not always possible in a cyberbullying context.

Fostering empathy is a key element of cyberbystander intervention programmes, see Henares-Montiel et al. (2023) for review. Trials of cyberbystander intervention programmes that aim to increase positive cyberbystander behaviour have called for more research to understand the components necessary to create effective programmes (DeSmet et al., 2018), as research is still lacking.   Furthermore, the review completed by Henares-Montiel et al. (2023) concluded the three intervention programs aimed at reducing passive cyberbystander behaviour were not effective. Therefore, there is a need to understand the mechanisms that would foster defending behaviours in cyberbystanders' to inform these interventions.

A systematic review completed by Zych et al. (2019) reported the results of studies that explored the relationship between empathy and different cyberbullying roles in children and adolescents. Cyberbullying roles investigated included cybervictims, cyberbullies and cyberdefenders. They found that cyberdefenders tended to score higher in empathy, but ultimately the authors concluded more studies were required to understand the strength of this relationship. In addition, they were unable to determine if a certain type of empathy (affective/cognitive) is more closely associated with cyberdefending or if empathy levels influence the other (passive, reinforcer) cyberbystander roles.

Since Zych et al's. (2019) review, more research has been completed on the relationship between empathy and cyberbystander behaviour when witnessing cyberbullying (Fabris et al., 2022; Hu et al., 2023; Leung &Chiu, 2023). Therefore, the present review is an update and extension of Zych et al's. (2019) review which aims to describe the quality, strengths and limitations of the current literature and clarify the relationship between empathy and cyberbystander behaviour.

Review Questions

1. What is the relationship between cognitive and/or affective empathy and cyberbystander behaviours in children and adolescents?
2. Is there evidence that suggests cognitive *or* affective empathy has more influence on cyberbystander behaviour?

Method

The present systematic review followed the PRISMA statement for guidance and structure throughout the process (Page at al., 2021). Please see Appendix A. for the PRISMA reporting checklist. A study protocol was uploaded to the Open Science Framework on the 5th of June 2024 which can be accessed via the following link: https://osf.io/43trm/?view_only=aae9814316fa402b90d16d50d8a11e5b .

Search Strategy

The framework below was used to define key elements of the research question.

1. **Population**: Children and adolescents. Children and adolescents were defined as young people aged 0-18 years (including 18-year-olds).

2. **Context**: Cyberbullying. Cyberbullying was defined as ' a set of behaviours performed through electronic or digital media by one individual or group of individuals who repeatedly communicate hostile or aggressive messages intended to inflict harm or discomfort on others' (Tokunaga, 2010).

3. **Phenomenon of interest:** Empathy. Distinguishing between cognitive and affective empathy. Studies measuring cognitive empathy defined as the ability to infer and understand others cognitive (e.g., thoughts, beliefs, intentions) and emotional states were included (Dorris et al., 2022).

Studies measuring affective empathy defined as the ability to share or understand the feelings of others which results in appropriate emotional responses were included (Reniers et al., 2011)

4. **Outcomes**: Bystander behaviour when witnessing cyberbullying. A recent systematic review characterised 2-5 bystander behaviours in children and adolescents (Polanco-Levicán & Salvo-Garrido, 2021). These broadly include passive, defending and reinforcing bystander behaviours. Studies which measured these bystander behaviours in a cyberbullying context were included.

The electronic databases searched in the review by Zych et al. (2019) were also searched in this review, PsycINFO (via EBESCO host); PubMed; SCOPUS; and Web of Science Core Collection (via Web of Knowledge). In addition, the Psychology and Behavioural Science Collection (via EBESCO host) was also searched. The previous review included Google Scholar searches, but in this review, Google Scholar was not searched because it is not reproducible.

The main search strategy was developed in consultation with a specialist Librarian at the University of Glasgow, supervisors and consideration of relevant published systematic reviews (Deng et al., 2021; Zych et al., 2019). Search terms for the present review, as used for PsycINFO are outlined below. The search strategy was amended slightly for each database (see appendix B1.).

Search Strategy for PsycINFO

1. DE "Cyberbullying"
2. TI (cyberbull* or cyber-bull*) OR AB (cyberbull* or cyber-bull*)
3. TI ((cyber or internet or digital or online or virtual or electronic) N2 (bull* or -bull* or harass* or agress* victim*)) OR AB ((cyber or internet or digital or online or virtual or electronic) N2 (bull* or -bull* or harass* or agress* victim*))
4. S1 OR S2 OR S3
5. Bystander Intervention
6. DE "Bystander Effect"
7. TI (bystand* or cyber-bystand* or cyberbystand* or cyber-def* or cyberdefend*) OR AB (bystand* or cyber-bystand* or cyberbystand* or cyber-def* or cyberdefend*)
8. S5 or S6 or s7
9. TI (child* or teen* or adolescen* or young* or school* or minor* or boy* or girl* or student* or youth* or juvenile* or kid* or underage* or preadol*) OR (child* or teen* or adolescen* or young* or school* or minor* or boy* or girl* or student* or youth* or juvenile* or kid* or underage* or preadol*)
10. S4 AND S8 AND S9

During initial scoping reviews it became apparent that in some studies which included several independent variables, empathy was sometimes omitted from the title and abstract. Therefore, to increase the retrieval of relevant studies, empathy was excluded from the search strategy. Therefore, studies that reported to measure cyberbystander behaviour and a range of factors sometimes referred to broadly as ‘psychological’ or ‘individual’ or ‘interpersonal traits’ were included, in case these studies measured empathy but did not explicitly state this in the title or abstract. These studies were then reviewed for the inclusion of empathy during full text screening. For example, Allison & Bussey’s (2017) paper titled ‘Individual and Collective Moral Influences on Intervention in Cyberbullying’ was included in the full text screening but was subsequently excluded because they did not measure empathy.

A date range was not applied as it was assumed cyberbullying being a relatively new issue, this would not be required. Age filters were not applied to the search results, as the factor of age was addressed through the search strategy.

Inclusion criteria:

- Quantitative research that focused on cyberbullying bystander behaviour and its association with cognitive and/ or affective empathy.
- Quantitative data from mixed methods studies if it could be extracted.
- Full text available.
- Written in English.
- Between subjects and correlational study designs.
- Longitudinal studies if cyberbystander behaviour and empathy were measured at the same time point.
- Intervention studies if there was data on the relationship between bystander behaviour in cyberbullying and empathy pre- intervention or in a control group.
- Includes data from children and adolescents aged 0-18 (up to and including age 18).

Exclusion criteria:

- Studies focusing on other types of online violence such as cyberstalking, cybertrolling and cybergrooming.
- Studies which focused explicitly on people with neurodevelopmental or psychiatric conditions.
- Studies that used 'imagined scenarios' to avoid tautological issues.
- Studies with definitions and conceptualisations that differed significantly from the definitions outlined.
- Qualitative research studies.
- Unpublished literature, systematic and literature reviews, case studies, conference abstracts or presentations, book sections and commentary pieces.
- Studies involving adults if the data for those aged 18 and younger could not be extracted separately.

In addition to the systematic search, manual searches were carried out using the reference lists from selected reviews (Hu et al., 2021; Polanco-Levicán & Salvo-Garrido, 2021; Zych et al., 2019) as well as forward and backward chaining from the final included studies. No further studies were identified.

Data Extraction and Synthesis of Findings

Data extraction was completed by the lead reviewer. Studies were extracted to Microsoft EndNote and duplicate studies were removed using the in-built tool as well as through methodical manualised sorting. The remaining studies after de-duplication were title and abstract screened. The remaining studies were then full text screened. Studies which met the inclusion criteria were quality assessed and the relevant data was extracted. A data extraction tool was developed for this review and can be seen below in Table 3. A narrative synthesis approach was applied to answer the review questions (Popay et al, 2006). The process of synthesising involved extracting key characteristics and reviewing key findings of included studies, describing and collating similar findings and identifying patterns in the evidence base for the relationship between empathy and cyberbystander behaviour. Where possible, outcomes were broadly compared using effect sizes.

Quality Assessment

The Crowe Critical Appraisal Tool (CCAT v1.4) (Crowe et al., 2013) was used to assess the quality of the included studies. Each of the included studies were scored on a scale from 0-5 across the following eight areas: preliminaries, introduction, design, sampling, data collection, ethical matters, results, and discussion. The CCAT was selected because it is suitable to use across a variety of research methodologies and provides detailed guidance.

The scores obtained from applying the CCAT to the studies, which are out of 40, were converted into percentages. Studies were then categorised according to their percentage score. Table 1 outlines the CCAT scoring key.

Table 1

CCAT Scoring Key

Quality Rating	Percentage	Equivalent CCAT Score
Poor Quality	≤49	19 or less
Acceptable Quality	**50 – 75**	**20-30**
High Quality	≥75	30-40

Inter-Rater Reliability

Inter-rater reliability was assessed at two stages: title and abstract eligibility and full-text screen reviewing. The second reviewer screened 10% of the titles and abstracts and 10% of the articles that were full text screened. As two reviewers were involved, Cohen's Kappa coefficients were calculated using an online calculator to assess inter-rater agreement. There was almost perfect agreement between the raters at the title and abstract screening stage (κ=.95). Perfect agreement was reached at the full screening review stage (κ=1.0). The second reviewer randomly selected two of the eight remaining studies to review for quality using the CCAT. There was a slight discrepancy on one of the domains of the CCAT which was resolved through conferencing.

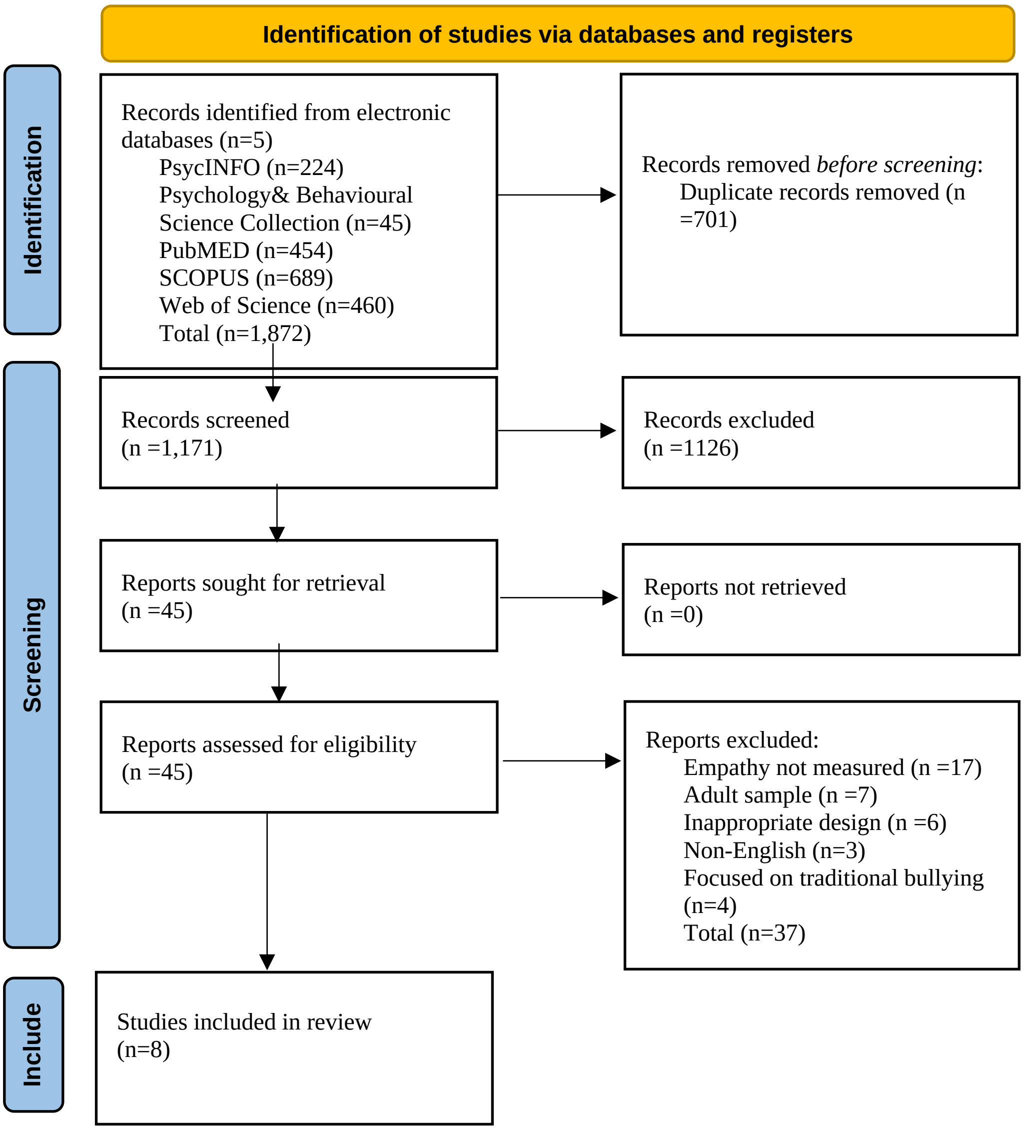
Results

Search Results

Final searches were conducted on 26th June 2024 and imported into EndNote for the removal of duplicates and for screening. A total of 1872 papers were identified. Following de-duplication, 701 articles were removed, resulting in 1171 unique references. Using the eligibility criteria, first the title and abstract were screened and papers which did not meet criteria were removed. Following abstract screening, 45 papers remained. Upon full text screening against the inclusion and exclusion criteria, a total of eight studies met eligibility criteria. All eight studies were retrieved. The PRISMA flow diagram below (figure 1) outlines this process.

Figure 1

PRISMA Diagram: Adapted from Page et al. (2021).



Quality Rating Results

Table 2 details the scores for each study across the 8 domains of the Crowe Critical Appraisal Tool (CCAT), as well as the total score, a transposed percentage, and the associated quality rating. All eight included studies were rated using the CCAT by the lead author. As shown in Table 2, three studies (Aliberti et al., 2022; DeSmet et al., 2016; Hu et al., 2023) were of ‘high quality’. Four studies were of ‘medium quality’ (Fabris et al., 2022; Leung et al., 2023; Machackova& Pfetsch, 2016; Van Cleemput et al., 2014), and one study was of ‘poor quality’ (Lui Zheng& Tung, 2018).

Methodological limitations were present in all studies in relation to the sampling design. None of the studies completed an a-priori calculation to establish the necessary sample size, nor did they provide justifications for why they did not complete one. Furthermore, none of the studies described why they selected their sample, convenience or otherwise. Five studies did not outline their inclusion and exclusion criteria (Fabris et al., 2022; Hu et al., 2023; Leung et al., 2023; Lui Zheng& Tung, 2018; Machackova & Pfetsch, 2016). Most of the studies lacked detail regarding recruitment and data collection procedures which would hinder replication except for Alberti et al. (2019). Apart from Alberti et al. (2019), and Hu et al. (2023) the remaining studies scored 3 or less on ethical issues. These studies did not address confidentiality, privacy or debrief procedures.

All the studies used appropriate statistical analyses for their research question(s). However, five of the studies did not include demographic descriptives (Fabris et al., 2022;

Leung et al., 2023; Lui Zheng& Tung, 2018; Machackova& Pfetsch, 2016; Van Cleemput et al., 2014)

The study completed by Lui Zheng& Tung (2018), scored poorly on all domains measured by the CCAT. Most notably, they did not include any statistics, tables or figures. They did not outline their recruitment, data collection or consent procedures. Therefore, the results derived from their study will not be drawn upon in the narrative synthesis.

A summary of quality rating scores is provided in Table 2.

Table 2.

CCAT Quality Ratings

Study	Preliminary	Introduction	Design	Sampling	Data Collection	Ethical Matters	Results	Discussion	Total Score (%)	Rating
Aliberti et al. (2022)	5	5	4	4	4	5	5	4	36(90%)	High
DeSmet et al. (2016)	4	5	5	4	5	3	5	5	36(90%)	High
Fabris et al. (2022)	4	5	5	3	2	3	3	4	29(73%)	Acceptable
Hu et al. (2023)	4	5	5	3	5	4	3	4	33(83%)	High
Leung et al. (2023)	4	5	4	3	3	3	4	3	29(73%)	Acceptable
Lui Zheng& Tung (2018)	1	2	1	2	0	0	1	1	8(23%)	Poor
Machackova& Pfetsch (2016)	4	5	3	3	3	2	4	4	28(70%)	Acceptable
Van Cleemput et al. (2014)	4	5	3	3	4	3	3	3	28(70%)	Acceptable

Study Characteristics

Of the eight studies included, three were completed in China (Hu et al., 2023; Leung et al.,2023; Lui Zheng & Tung, 2018), two were completed in Germany (Machackova & Pfetsch, 2016; Van Cleemput et al., 2014), one in the United States (Aliberti et al., 2022), one in Belgium (DeSmet et al., 2016), and one in Italy (Fabris et al., 2022).

There was variability in the overall sample sizes which ranged from 321-2,333 participants. The age of participants varied in the studies but ranged from 9-18 years. Most studies recruited participants from school (Alberti et al., 2022; DeSmet et al., 2016; Hu et al., 2023; Leung et al., 2023; Lui Zheng& Tung, 2018).

All the studies were cross-sectional in design with the analysed data collected from one time point. The data utilised in Machackova & Pfetsch, (2016) study was part of a wider longitudinal study. Five studies measured both cognitive and affective empathy (Aliberti et al., 2022; Fabris et al., 2022; Hu et al., 2023; Leung et al., 2023; Machackova & Pfetsch, 2016). Although Hu et al. (2023) did not differentiate between cognitive or affective empathy despite using a scale that measures both domains (The Basic Empathy Scale). Two studies measured affective empathy only (DeSmet et al., 2016; Van Cleemput et al., 2014) and one measured cognitive empathy only (Lui Zheng & Tung, 2018).

All studies employed standardised self-report measures of empathy. Three studies administered the Basic Empathy Scale (Hu et al., 2023; Leung et al., 2023; Machackova &

Pfetsch, 2016). Three studies administered the Interpersonal Reactivity Index (DeSmet et al., 2016; Fabris et al., 2022; Lui Zheng & Tung, 2018). One study administered the Empathic Responsiveness Scale (Van Cleemput et al., 2014) and one study administered the How I Feel in Different Scenarios Scale (Aliberti et al., 2022). Two studies compared male and female scores (Aliberti et al., 2022; Hu et al., 2023).

Six studies reported results for defending cyberbystander behaviour only (Aliberti et al., 2022; DeSmet et al., 2016; Hu et al., 2023; Leung et al., 2023; Lui Zheng & Tung, 2018, Van Cleemput et al., 2014). One study reported differences in cognitive and affective empathy between passive and defending cyberbystanders (Fabris et al., 2022). Lastly, Machackova & Pfetsch (2016) measured the relationship between cognitive and affective empathy and reinforcing and defending cyberbystander behaviour.

A summary of the included studies characteristics is provided in table 3.

Table 3

Data Extraction: Summary of the 8 Included Studies that Met Criteria

Authors and Location	**Number of Participants and age/grade**	**Study Design**	**Cyberbystander Role(s) Measured**	**Measures of Cyberbystander Behaviour and Empathy**	**Analysis**	**Relevant Findings**
Aliberti et al. (2022) United States	872 6^{th}-12^{th} grade students. (11–18-year-olds)	Cross-sectional	Intervening	The adapted version of the 16-item Bystander Intervention in Bullying (Jenkins et al., 2018) was adopted for cyberbullying. Three items related to active intervening. How I Feel in Different Scenarios Scale (Empathy- cognitive and affective)	Bivariate Correlations	AE was statistically significantly positively correlated with helping in males (r=.18, p<.01) and females (r=.32, p<.01) CE was statistically significantly positively correlated with helping in males (r=.30, p<.01) and females (r=.24, p<.01)
					Structural equation modelling (SEM) Multi-group Path Analysis	AE was positively and significantly related to helping in females only (females β = .237, p < .001) (though chi-square difference was not significant) CE was not significantly related to helping in females (β=.055, p=.397) or males (β=.146, p=.046)
DeSmet et al. (2016)	1979 7^{th}-9^{th} grade students (16	Cross-sectional	Defending	Ten items measured specific bystander reactions (yes/no) and behavioural intentions	Bivariate Correlation	Defending bystander behaviour was statistically

Belgium	schools) (12-15 years)			on a 5-point scale. The content of these items was based on prior qualitative research (DeSmet et al., 2014). Interpersonal Reactivity Index (Affective Empathy only)		significantly related to AE (*r*=.19, p < .01)
Fabris et al. (2022) Italy	1158 children and adolescents. (11-15 years)	Cross-sectional	Uninvolved students (N = 370), defending (N = 258), passive bystanders (N = 88), and both defender and passive bystander (442).	Bullying and Observer Behaviours in Cyberbullying Episodes (Pozzoli & Gini,2020). 16 item scale. 4 items related to defending and 4 items to acting passively. Interpersonal Reactivity Index (IRI) (Cognitive and affective empathy)	MANCOVA	Controlling for age and sex. Significant differences between the bystander groups CE scores (p<.01, η^2= .02) Defending bystanders had statistically significant (higher) CE scores compared to the other types of bystander groups. There were no differences between the other bystander groups CE scores. No difference in AE.
Hu et al. (2023). China	919 (812)† 7th-8th graders (12-14).	Cross-sectional	Defending	Styles of Bystander Intervention Scale (Moxey & Busy, 2019) 15 item scale focusing on constructive interventions e.g. encouraging victim to report or aggressive interventions e.g. threatening the bully. Basic Empathy Scale (Cognitive and affective).	Bivariate correlations Structural equation modelling (SEM) PROCESS marco (mediation moderation)	Bivariate correlations showed that empathy was significantly positively associated with bystander helping behaviour in cyberbullying (r= .18 p < .001) in males and females. Even after considering the indirect effect through internet moral judgment, the direct effect of empathy on bystander helping behaviour remained significant (β= .14, p < .001).

Leung et al. (2023) China	817 secondary school students (M age-15.1)	Cross-sectional	Defending	Subscale adapted from the Participant Role Questionnaire (PRQ) measured cyberdefending behaviour This scale encompasses 3 items for three cyberbystander roles: reinforcer, defender and passive Basic Empathy Scale (Cognitive and affective empathy).	Bivariate Correlation Structural equation modelling (SEM)	There was a statistically significant relationship between cyber-defending and AE (r=.11, p=<.01). There was no correlation coefficient for CE. Participants with 10% higher AE than the mean averaged 0.5% more cyber defending, mediated by 3.8% higher pro-victim belief and 1.7% higher cyberbullying awareness. CE did not influence cyberbystander behaviour
Lui Zheng& Tung (2018) China	884 7-9th grade students (9-12) ‡	Cross-sectional	Defending	The bystander behavioural intentions questionnaire (no further information provided) Interpersonal Reactivity Index (Cognitive Empathy)	Structural equation modelling (SEM) Multi-group analysis	CE was positively associated with cyber-defending. No statistics were reported
Machackova & Pfetsch(2016)	321 adolescents 12-18 years.	Cross-sectional	Defending and reinforcing	Bystander response to cyberbullying was measured with a German scale adapted from Steffgen, Happ and Pfetsch (2013). Three items related to supportive responses to victim of cyberbullying, for example, “I tried to stop bullying via Facebook and/or YouTube.”). Three statements were related to the reinforcement of the bully in in cyberbullying, for example, “I joined in humiliating some classmates via Facebook’).	Bivariate Correlation Path model with observed variables (Mplus 6.1, MLR estimator)	CE (r =.21, p < .01) and AE (r=.26, p < .01) were both positively related to defending the victim of cyberbullying. Both CE (r=-.07) and AE (r=-.04) were negatively correlated with reinforcement of the cyberbully, but this was not statistically significant. Structural equation path model including gender, age, normative beliefs about verbal and cyberaggression showed that AE (β=.19, p < .05) was a

				Basic Empathy Scale (Cognitive and affective)		significant individual predictor of defending the victim whereas CE was not a significant individual predictor of defending (β not reported)
Van Cleemput et al. (2014) Germany	2,333 children and adolescents (9–16)	Cross-sectional	Defending	Questions relating to actions when witnessing cyberbullying (non-standardised) Empathic Responsiveness Scale (modified version) (Affective empathy only)	Bivariate Correlation Structural equation modelling (SEM)	AE was positively related to defending the victim ($r=.35$, $p < .01$). A structural equation model including gender, age, cyber-perpetration, cyber-victimisation, face-to-face victimisation and perpetration and social anxiety showed that empathy was an independent predictor of defending ($β=.42$, $p < .01$)

Abbreviations: AE, Affective Empathy; CE, Cognitive Empathy.

†: This is the number of valid data included in the analysis.

‡: Age range not provided by authors. The age range of 9-12 is approximate based on school grade provided.

Narrative Synthesis of Results

Question 1: What is the relationship between cognitive and/or affective empathy and cyberbystander behaviour?

Six studies (Aliberti et al., 2022; DeSmet et al., 2016; Hu et al., 2023; Leung et al., 2023; Machackova & Pfetsch, 2016; Van Cleemput et al., 2014) reported correlation coefficients for the relationship between empathy and cyberbystander defending behaviour. There was variation in the type of empathy studied but all reported statistically significant positive, but weak, correlation coefficients ranging from .18 to .35 suggesting those higher in empathy engage in more cyberdefending behaviour when they witness cyberbullying.

In addition to defending behaviour, Machackova & Pfetsch (2016) measured reinforcing cyberbystander behaviour. As outlined in table 3, reinforcer behaviour was measured using an adapted scale from Steffgen et al. (2013). Three items related to the reinforcement of the cyberbully. An example of a sample item was, "I joined in humiliating some classmates via Facebook'. This study reported cognitive and affective empathy negatively correlated with reinforcing behaviour, however these were non-significant correlations ($r=-.04$) for cognitive empathy and ($r=-.07$) for affective empathy.

Using structural equation modelling, three studies reported affective empathy remained a significant predictor of cyberdefending when controlling for the influence of other variables (Alberti et al., 2022; Machackova & Pfetsch; 2016; Van Cleemput et al., 2014). Hu et al. (2023) did not differentiate between types of empathy but reported, the direct effect of empathy on bystander helping behaviour remained significant ($\beta = .14$, $p < .001$) after controlling for the indirect effect of internet moral judgment. The other structural equation modelling results will be described under question 2.

Question 2: Is there any evidence that suggests cognitive or affective empathy has more influence on cyber-bystander behaviour

Considering the reported r values (correlation coefficients) as effect sizes in statistically significant bivariate correlation results, the results do not provide unequivocal evidence that cognitive or affective empathy is more influential on cyberbystander behaviour.

Four studies differentiated between cognitive and affective empathy (Aliberti et al., 2022; Fabris et al., 2022; Leung et al., 2023; Machackova & Pfetsch, 2016). Two studies (Aliberti et al., 2022; Machackova & Pfetsch, 2016) reported both types of empathy were statistically significantly associated with cyberdefending behaviour. However, Leung et al. (2023) reported cognitive empathy had no influence on cyberdefending behaviour, but affective empathy did in bivariate correlations. Leung et al's. (2023) structural equation model demonstrated affective empathy indirectly explained cyberdefending through greater pro-victim beliefs and greater awareness of cyberbullying. Meaning, people with higher affective empathy are more likely to appreciate victims suffering which prompts defending behaviour, compared to cognitive empathy which was found to not be directly related to defending behaviour.

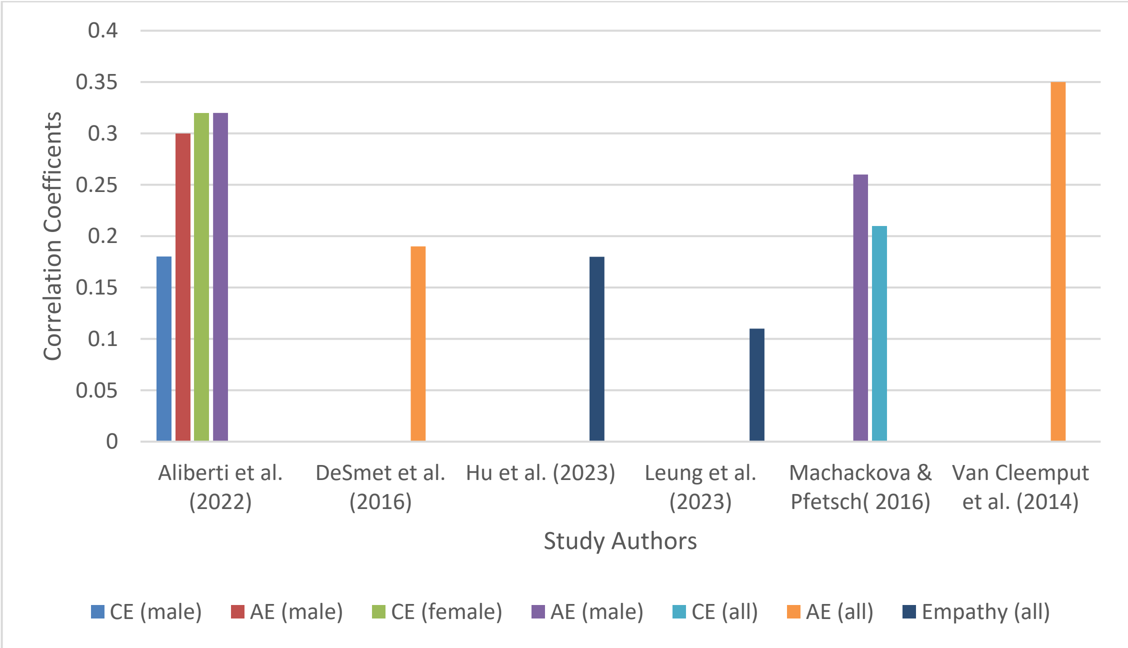
Alberti et al. (2022) and Machackova & Pfetsch, (2016) also completed further structural equation modelling. Alberti et al. (2022) reported affective empathy remained significantly related to cyberdefending in females only ($\beta = .237$, $p < .001$), and the effect of cognitive empathy weakened and was no longer statistically significant. Although chi-square difference between males and females was not significant. Similarly, Machackova & Pfetsch, (2016) completed a structural equation path model including gender, age, normative beliefs about verbal and cyberaggression and found that affective empathy remained a significant individual predictor of defending the victim whereas cognitive empathy did not. Van Cleemput et al. (2014) only measured affective empathy but also reported affective empathy remained a direct predictor of defending bystander behaviour when included in a structural equation model with gender, age, cyber-perpetration, cyber-victimisation, face-to-face victimisation and perpetration and social anxiety.

Fabris et al. (2022) measured differences between cyberdefenders and those who remain passive when witnessing cyberbullying cognitive and affective empathy levels. They reported significant differences between the group's cognitive empathy scores ($p<.01$, $\eta2= .02$). Defending cyberbystanders had statistically significant (higher) cognitive empathy scores compared to passive cyberbystanders, but there were no differences in affective empathy scores between the bystander groups.

Figure 2 summarises the results of the synthesis in relation to empathy and cyberbystander defending behaviour for the studies which provided correlation coefficients.

Figure 2.

Correlation Coefficients for Cyberbystander Defending Behaviour and Empathy



Abbreviations: AE, Affective Empathy; CE, Cognitive Empathy

Discussion

The present review demonstrates that research on cyberbystander behaviour has continued to grow and develop; four of the eight included studies were published between 2022 and 2023. However, compared to bystander behaviour in traditional bullying, the research landscape is still lacking.

In relation to the range of cyberbystander behaviours; defending, reinforcing and passive, only one study measured the relationship between empathy and reinforcing cyberbystander behaviour (Machackova & Pfetsch, 2016), who reported a non-significant correlation. The remaining correlational studies focused on cyberdefending bystander behaviour only. Therefore, due to the lack of studies on other cyberbystander behaviours of interest (reinforcing and passive), the current review cannot draw conclusions on the relationship between those behaviours and empathy.

The studies that described the relationship between empathy and defending cyberbystander behaviour found statistically significant but weak correlations. The systematic review and meta-analysis completed by Deng et al. (2021) which focused on bystander behaviour and empathy in traditional bullying, reported an average correlation coefficient of ($r=.27$) for affective empathy and ($r=.22$) for cognitive empathy. While these are weak correlations, the similarity between these results, and the correlations of the studies included in the present systematic review indicate that, despite the differences between online and traditional bullying, the strength of the relationship between bystander defending and empathy is similar in both cyber and traditional bullying contexts.

The present systematic review does not have enough evidence to conclude if affective or cognitive empathy is more influential on cyberdefender behaviour. Only one study of acceptable quality reported cognitive empathy was not statistically significantly correlated with cyberdefending (Leung et al., 2023). However, only affective empathy remained a significant predictor of cyberdefending behaviour in the studies that differentiated between cognitive and affective empathy when the effect of other variables was accounted for during structural equation modelling analysis (Alberti et al. 2022) (females only), (Machackova & Pfetsch, 2016),

These findings may be considered alongside research on school bullying. Research has reported cognitive empathy was significantly positively correlated with three stages that occur prior to defending behaviour, namely, paying attention to a school bullying incident, undertaking intervention responsibility, and knowing how to deal with a bullying incident. Whereas affective empathy was significantly positively correlated with interpreting the situation as an emergency and the actual defending behaviour (Fredrick et al., 2020). These results indicate that cognitive empathy is important in the initial stages of defending, whereas affective empathy is necessary to spur participation in defending behaviour. However, in Alberti et al's, (2022) study, which was included in this review, results from their multi-group path analysis, demonstrated cognitive empathy was not associated with any of the steps of bystanding which included noticing, interpreting, accepting responsibility, knowing how to help and intervening. The authors suggested anonymity and reduced social cues may be responsible for the lack of cognitive empathy activation in an online situation compared to an offline one.

While it tentatively appears, affective empathy could have a more direct influence on the *behavioural* aspect of cyberdefending, and cognitive empathy could be more influential in the initial stages of bystander behaviour; interpreting, noticing and accepting responsibility, further research is required. Research underpinned by models of bystander behaviour, for example 'The Bystander Intervention Model' (Latané & Darley, 1970), which was utilised in Alberti et al's. (2022) study may help clarify this. These outcomes may have important implications for the development of cyberbullying intervention and reduction programs.

Lastly, sex differences between the influences of empathy on cyberbystander defending behaviour were only reported by Alberti et al. (2022). They reported affective empathy only remained a significant predictor of defending in females following structural equation modelling, although significant sex differences were not found in the follow-up chi-square analysis. This suggests that while there is difference between males and females, it is not a significant difference. Both Machackova & Pfetsch, (2016) and Van Cleemput et al. (2014) controlled for sex in their structural equation model and still found affective empathy to be a direct predictor of cyber defender behaviour which suggests the relationship persists even after considering the impact of sex. The literature on sex differences in cyberbystanding behaviours is very mixed with some reporting females tend to engage in more helping behaviours while others state there is no difference between males and females (Kozubal et al., 2019). Further research would be helpful to ascertain if biological sex moderates the influence of empathy on cyberdefending behaviour.

Review Limitations and Strengths

The studies included in the present systematic review have similar methodological flaws. All the studies utilised self-report measures which are subject to social desirability effects and lack ecological validity particularly as children and adolescents do not have fully matured empathy or self-awareness abilities. Although, it did mean the results of the studies included in the review were more comparable due to the same collection methods. Furthermore, all the studies were cross-sectional and therefore no causal relationships can be established based on its results. Most of the studies utilised school recruitment which is suitable based on the target population and likely increased the opportunity for children and adolescents from communities that typically do not engage in research or find it difficult to access research to engage. However, the use of school recruitment may bias the results if schools from particular social-demographic areas typically always engage in research. While school sampling will likely provide a diverse and large sample size, future research studies could be strengthened by recruiting from various venues and engaging in outreach.

In relation to limitations of the specific review, the review is subject to publication bias, in that studies with significant findings are more likely to be published and therefore included in this review. With that in mind, all the studies included in the review reported significant results. Furthermore, the findings of the review are not generalisable as the studies included only came from 5 different countries, despite cyberbullying occurring in at least 63 countries (Zhu et al., 2021). Furthermore, 62% of the studies included in this review were either of acceptable or poor quality. One of the studies was felt to be too methodologically flawed and did not provide enough information for their results to be included in the narrative synthesis (Lui & Zheng, 2018). The results drawn from the review could be strengthened by a meta-analysis and more research studies. The use of a narrative, rather than a more structured synthesis, may be considered a limitation of this systematic review.

Strengths of the present review include the exclusion of empathy from the search strategy and use of the CCAT. The CCAT is also accompanied by user guidance which adds to its validity through increasing the uniformity of its application. Rating pairs were also used in the present review at three stages: title and abstract screening, full text screening and during quality rating. This strengthens the present review through the reduction of subjectivity bias, errors, and increasing transparency.

Future Research

This study searched five databases and used a sensitive search strategy to retrieve as many relevant articles as possible. The inclusion of only 8 studies demonstrates that the research landscape is lacking, particularly in relation to passive and reinforcing cyberbystander roles. Furthermore, research studies from only 5 countries were included. Therefore, further research underpinned by theory is required to assess if the results reported here are replicated in other countries and in more studies of varying sample size.

The research landscape would benefit from longitudinal studies that measure how the relationship between empathy and cyberbullying bystander behaviour changes over time, particularly as research suggests social cognitive abilities continue to develop, not always linearly, throughout adolescence (Dorris et al., 2022). Understanding this interaction is important for the development of cyberbullying reduction programmes. To increase ecological validity, an empirical experiment where participant's natural responses to observing cyberbullying would be ideal, however, careful consideration of ethical issues would be vital. Lastly, many of the studies adapted scales developed for use in traditional bullying research for use in their cyberbullying research. Furthermore, these scales usually included only a small number of items. Future research would benefit from utilising scales that have now been standardised and created to measure bystander behaviour in cyberbullying specifically such as the 16-item scale developed by Pozzoli et al. (2019) which was utilised in Fabris et al's. (2022) study.

Given that the relationship between empathy and cyberbystander behaviour is weak, future research should focus on other factors that promote pro-social cyberbystander behaviour in children and adolescents. A recent literature review reviewed the evidence for the individual, contextual and social environmental factors that influence cyber bystander responses (Machackova, 2020). The author noted researchers need to pay more attention to the contextual factors that differ from offline bullying that influence pro-social behaviour. Contextual factors can include language used by the cyberbully or the platform used, which help form judgments about the severity of aggressive incidents in the online environment, which in turn influence the cyberbystanders' decision to provide help. These cues, if very different from traditional bullying, likely impact the activation of individual factors including empathy.

Conclusion

This review highlighted there is still a need for further research to be completed into the relationship between empathy and cyberbystander behaviours, particularly passive and reinforcing behaviours. The research landscape would benefit from studies being completed in more countries and with measures specific to cyberbystander behaviour. It does appear that there is enough evidence to confirm a significant positive, but weak relationship between empathy and cyber defending behaviour. There is early suggestion that affective empathy has a direct influence on defending behaviour when other variables are controlled for compared to cognitive empathy, but further research is required.

There were no sources of funding for this project or competing interests.

References

Aliberti, M., Jenkins, L., & Monk, M. (2023). Predictors of Cyberbystander intervention among adolescents. *Journal of interpersonal violence*, *38*(9-10), 6454-6479.

Allison, K. R., & Bussey, K. (2016). Cyber-bystanding in context: A review of the literature on witnesses' responses to cyberbullying. *Children and Youth Services Review*, *65,* 183-194.

Allison, K. R., & Bussey, K. (2017). Individual and collective moral influences on intervention in cyberbullying. Computers in Human Behavior, 74, 7-15.

Aparisi, D., Delgado, B., Bo, R. M., & Martínez-Monteagudo, M. C. (2021). Relationship between cyberbullying, motivation and learning strategies, academic performance, and the ability to adapt to university. *International journal of environmental research and public health*, *18*(20), 10646.

Bronfenbrenner, U. (2000). *Ecological systems theory*. American Psychological Association.

Cohen's kappa free calculator. (n.d.). https://meta-analysis.actilab.onl/cohen-kappa-free-calculator/

Crowe, M., Crowe critical appraisal tool (CCAT) user guide. Conchra House, 2013. 10.

Deng, X., Yang, J., & Wu, Y. (2021). Adolescent empathy influences bystander defending in school bullying: A three-level meta-analysis. Frontiers in psychology, 12, 690898.

DeSmet, A., Bastiaensens, S., Van Cleemput, K., Poels, K., Vandebosch, H., Deboutte, G., ... & De Bourdeaudhuij, I. (2018). The efficacy of the Friendly Attac serious digital game to promote prosocial bystander behavior in cyberbullying among young adolescents: A cluster-randomized controlled trial. *Computers in Human Behavior*, *78*, 336-347.

DeSmet, A., Bastiaensens, S., Van Cleemput, K., Poels, K., Vandebosch, H., Cardon, G., & De Bourdeaudhuij, I. (2016). Deciding whether to look after them, to like it, or leave it: A multidimensional analysis of predictors of positive and negative bystander behavior in cyberbullying among adolescents. *Computers in Human Behavior*, *57*, 398-415.

Dorris, L., Young, D., Barlow, J., Byrne, K., & Hoyle, R. (2022). Cognitive empathy across the lifespan. *Developmental Medicine & Child Neurology*, *64*(12), 1524-1531.

Fabris, M. A., Longobardi, C., Morese, R., & Marengo, D. (2022). Exploring multivariate profiles of psychological distress and empathy in early adolescent victims, bullies, and bystanders involved in cyberbullying episodes. *International journal of environmental research and public health*, *19*(16), 9871.

Fredrick, S. S., Jenkins, L. N., & Ray, K. (2020). Dimensions of empathy and bystander intervention in bullying in elementary school. *Journal of school psychology*, *79*, 31-42.

Henares-Montiel, J., Pastor-Moreno, G., Ramírez-Saiz, A., Rodriguez-Gomez, M., & Ruiz-Pérez, I. (2023). Characteristics and effectiveness of interventions to reduce cyberbullying: a systematic review. Frontiers in Public Health, 11, 1219727.

Hong, J. S., Zhang, S., Wright, M. F., & Wachs, S. (2023). Racial and ethnic differences in the antecedents of cyberbullying victimization in early adolescence: An ecological systems framework. *The Journal of Early Adolescence*, *43*(1), 59-89.

Hu, Y., Bai, Y., Pan, Y., & Li, S. (2021). Cyberbullying victimization and depression among adolescents: A meta-analysis. *Psychiatry research*, *305*, 114198.

Hu, Yang, Tian Zhang, Hui-fen Shi, and Cui-ying Fan. "Empathy and bystander helping behavior in cyberbullying among adolescents: the mediating role of internet moral judgment and the moderating role of internet self-efficacy." *Frontiers in Psychology* 14 (2023): 1196571.

Huang, X., Chu, X., Liu, Q., Zhou, Z., & Fan, C. (2019). Bystander behavior in cyberbullying. *Advances in Psychological Science*, *27*(7), 1248.

Jenkins, L. N., Fredrick, S. S., & Nickerson, A. (2018). The assessment of bystander intervention in bullying: Examining measurement invariance across gender. *Journal of school psychology*, *69*, 73-83.

Jiang, Q., Zhao, F., Xie, X., Wang, X., Nie, J., Lei, L., & Wang, P. (2022). Difficulties in emotion regulation and cyberbullying among Chinese adolescents: A mediation model of loneliness and depression. *Journal of interpersonal violence*, *37*(1-2), NP1105-NP1124.

Kozubal, M., Szuster, A., & Barlińska, J. (2019). Cyberbystanders, affective empathy and social norms. Studia Psychologica, 61(2), 120-131.

Latané, B., & Darley, J. M. (1970). The unresponsive bystander: Why doesn't he help?. (No Title).

Leung, A. N. M., & Chiu, M. M. (2023). Adolescents' cyber-defending for cyberbullying: a socio-emotional, beliefs, and past experience model. *Cyberpsychology, Behavior, and Social Networking*, *26*(10), 789-797.

Li, C., Wang, P., Martin-Moratinos, M., Bella-Fernandez, M., & Blasco-Fontecilla, H. (2022). Traditional bullying and cyberbullying in the digital age and its associated mental health problems in children and adolescents: a meta-analysis. *European child & adolescent psychiatry*, 1-15.

Liu, Z. J., & Tung, Y. (2018). The impact of peer relationships, conformity, and cognitive empathy of bystander behavioral intentions for cyberbullying. *Chinese Journal of Psychology, 60*(2), 101-124.

Machackova, H. (2020). Bystander reactions to cyberbullying and cyberaggression: individual, contextual, and social factors. Current opinion in psychology, 36, 130-134.

Machackova, H., & Pfetsch, J. (2016). Bystanders' responses to offline bullying and cyberbullying: The role of empathy and normative beliefs about aggression. *Scandinavian journal of psychology*, *57*(2), 169-176.

Moxey, N., & Bussey, K. (2020). Styles of bystander intervention in cyberbullying incidents. International journal of bullying prevention, 2(1), 6-15.

Obermaier, M., Fawzi, N., & Koch, T. (2016). Bystanding or standing by? How the number of bystanders affects the intention to intervene in cyberbullying. *New media & society*, *18*(8), 1491-1507.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *bmj*, *372*.

Polanco-Levicán, K., & Salvo-Garrido, S. (2021). Bystander roles in cyberbullying: A mini-review of who, how many, and why. Frontiers in psychology, 12, 676787.

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., & Duffy, S. (2006). Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme Version, 1(1), b92.

Pozzoli, T., & Gini, G. (2020). Behavior during cyberbullying episodes: Initial validation of a new self-report scale. Scandinavian journal of psychology, 61(1), 22-29.

Reniers, R. L., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. A. (2011). The QCAE: A questionnaire of cognitive and affective empathy. *Journal of personality assessment*, *93*(1), 84-95.

Song, J., & Oh, I. (2018). Factors influencing bystanders' behavioral reactions in cyberbullying situations. *Computers in Human Behavior*, *78*, 273-282.

Steffgen, G., Happ, C., & Pfetsch, J. (2013). Measuring bystander bahaviour in bullying and cyberbullying incidents.

Tokunaga, R. S. (2010). Following you home from school: A critical review and synthesis of research on cyberbullying victimization. Computers in human behavior, 26(3), 277-287.

Zych, I., Baldry, A. C., Farrington, D. P., & Llorent, V. J. (2019). Are children involved in cyberbullying low on empathy? A systematic review and meta-analysis of research on empathy versus different cyberbullying roles. *Aggression and violent behavior*, *45*, 83-97.

Chapter 2

Social Responsiveness in 9-12 Year Olds: An ABCD Cohort Study.

Prepared in accordance with the author requirements for British Journal of Psychology.
https://bpspsychub.onlinelibrary.wiley.com/hub/journal/20448295/homepage/forauthors.html

Plain Language Summary

Title

Social Responsiveness in 9-12 Year Olds: An ABCD Cohort Study.

Background

Adaptive social interactions rely on a range of social cognitive abilities, for example being able to understand other's perspectives. People with neuropsychiatric or neurodevelopmental conditions such as autism sometimes experience social difficulties due to differences in the development of their social cognitive abilities. However, these difficulties and differences can also be observed in people without these conditions. This can result in 'normal differences' being pathologised. Therefore, there is interest in research aimed at understanding social cognition in children and adolescents with and without these conditions. Most social developmental research reports a female advantage, but further research is required to understand the mechanisms behind this advantage.

Social responsiveness (SR) is defined as the ability to engage appropriately in social interactions, encompassing the ability to interpret and respond to social cues, navigate social conventions, and adapt behaviour according to situational demands (Constantino & Gruber, 2012). This study aimed to understand how SR presents in children aged 9-12 years and if this is similar in males and females. The second aim was to determine to what extent socioeconomic status (SES) and performance on a range of cognitive tasks were associated with SR.

Methods

Data was taken from the ‘Adolescent Brain Cognitive Development study’. This is a longitudinal study based in the United States that recruited 11,875 children aged 9-10 between 2016-2018. Participants with a current diagnosis of schizophrenia, moderate/severe autism, intellectual disability and alcohol/substance use were not recruited to the ABCD study. In this study, we included 9,804 unrelated participants. SR was measured using a shortened version of the ‘Social Responsiveness Scale’ (SRS). Cognitive abilities were measured by performance on a range of cognitive tasks. The data was analysed using statistical software.

Main Findings

Between ages 9-12 years there was not a significant change in SR ability. However, there was a significant difference between male and females SR scores at age 9, 10, 11 and 12 due to slightly higher scores (indicating poorer SR) among males. SES and some of the cognitive abilities were significant predictors of SR, but overall, these factors have very little influence on SR. This indicates there are other factors not included in this study that would better explain SR ability and the mechanisms behind the female advantage.

Conclusion

The results aligned with previous research which show a female advantage in social cognitive abilities. While SES and some of the cognitive factors were significantly associated with SR ability, these do not hold much practical or clinical importance. Future research should focus on understanding the other factors that may influence SR including biological theories such as hormone exposure in-utero and social theories around environmental influences to add to our clinical and scientific understanding of social cognitive development and ability.

References

Constantino, J. N., & Gruber, C. P. (2012). *Social responsiveness scale second edition (SRS2): Manual*. Western Psychological Services (WPS).

Abstract

This study reports on the social responsiveness (SR) of 9,804 9–12-year-olds from the Adolescent Brain Cognitive Development study. SR was defined as the ability to engage appropriately in social interactions as measured by the Social Responsiveness Scale (SRS). Between groups analysis examined age and sex differences in SR ability. Regression analyses assessed the association between SR and cognitive test scores and socioeconomic factors. There was a significant difference between male and female SR scores F (1, 9213) =30.2, p=.001) with a small effect size (η^2_p=.003), with males scoring higher than females at all ages. The cognitive scores accounted for little variance in SR. Future research, study strengths and limitations are discussed. Alternative explanations for the female advantage including biological theories that may influence SR development such as in-utero androgen exposure and social theories around environmental shaping of gender-based roles are described.

Introduction

Social behaviour develops across the lifespan and is initially supported by foundational cognitive abilities in infancy such as facial processing and joint attention. Neonates display a visual attention preference for human faces (Reynolds & Roth, 2018) and seek eye contact and mirror the expressions of their caregivers to seek attunement and interaction (Tronick et al., 1978). By age 3-4, most children have an understanding that others can hold different and false beliefs to their own (Hofmann et al., 2016) and use words such as 'happy' and 'sad' to describe others internal experiences (Wellman, 2014). More complex Theory of Mind (ToM) abilities such as 'double bluffs' appear to develop in early childhood and adolescence (8-13 years) (Devine & Hughes, 2013).

Most research on the development of social cognitive abilities has been completed with very young children or with people with neurodevelopmental or mental health conditions. This is understandable as people with these conditions can have social communication difficulties because of differences in the development of their social cognitive abilities (see Fatima & Babu, 2023 for a systematic review). However, differences in social abilities because of social cognition differences observed in people with autism are not exclusive to this condition (Lyall et al., 2014). Indeed, there has been a recent tendency to view 'autistic traits' as being observable on a continuum in the general population (Constantino, 2021). Moreover, autism screening tools commonly used in clinical practice such as the Social Responsiveness Scale (SRS) (Constantino & Gruber, 2012) and measures of specific social cognitive abilities, for example the Reading the Mind in the Eyes test

(RMET) (Baron-Cohen et al., 2001) have been validated and have utility to capture the broader social cognitive phenotype in the general population.

A recent study by Dorris et al. (2022), explored cognitive empathy ability in 4545 participants aged less than 5 years to more than 75 years old across the lifespan. The study utilised the Reading the Mind in the Eyes Test (RMET) (Baron-Cohen et al., 2001). Dorris et al. (2022) reported that the 508 adolescents aged 13 to 18 scored on average 10% lower, indicating poorer performance, than the 543 children aged 10 to 12. This poorer performance in cognitive empathy was observed in both male and female participants aged 13-18, despite females having enhanced cognitive empathy abilities in this study overall. This finding suggests that the development of cognitive empathy may not be linear in typically developing young people. Further research measuring different aspects of social development is required to investigate this pattern further and if the pattern is consistent in both males and females. A commentary on the study by Dorris et al. (2022), by Baron-Cohen and colleagues, (2022), also emphasised the need for further research to understand the female advantage, and the mechanisms behind the advantage.

A female sex advantage in social cognitive abilities was reported in a recent systematic review and meta-analysis (Wood-Downie et al., 2021). This review included studies which utilised the SRS and reported typically developing females had significantly better social communication and interaction skills compared to the males across the included studies. There is ongoing debate as to whether the female advantage is fixed and stable due to biological differences or whether socio-cultural influences shape these abilities (Miller

&Halpern, 2014). Therefore, more research is required to understand the female advantage, when it emerges and the mechanisms behind the advantage.

The SRS (Constantino & Gruber, 2012) is a parent report measure that can be used to assess social abilities in 4–18-year-olds with and without autism. Social responsiveness (SR) refers to an individual's ability to engage appropriately in social interactions, encompassing the ability to interpret and respond to social cues, navigate social conventions, and adapt behaviour according to situational demands (Constantino & Gruber, 2012). Studies using the SRS have reported age and sex differences in SR scores in typically developing young people. Wallace et al. (2017) reported 438 children aged 4-8, 9-11 and 12-17 years were rated as having significantly fewer social impairments with increasing age. Some research reported no sex differences in SRS scores (Backer van Ommeren et al., 2017; Wallace et al., 2017; Sedgewick et al., 2019). However, other research has found that male sex is associated with higher (poorer) SRS scores (Hus et al., 2013). SRS standardisation samples also reported higher SRS scores among males than females on parent rated forms (Constantino & Gruber, 2012). What is less clear from the research is if the female advantage is consistent across the age range or if the sex differences in social responsiveness is different depending on age.

In addition to biological sex and age, several theoretical models explain how environmental, cognitive, and biological factors interact to support social competence (see Beaudoin & Beauchamp, 2020 for review). Despite conceptual differences, there is agreement that the development and maintenance of social skills is a complex and interactional process. Correlational and longitudinal studies have reported associations between higher parental income and/or education and increased pro-social behaviour in children, though some studies found no association or contrary results (Silke et al., 2018). Some authors have suggested that this advantage may be because higher income families have more resources to support their children's physical and emotional development (Matthews & Gallo, 2011). Other research suggests economic stress may reduce or deplete young people's emotional and cognitive resources to engage and connect with others (Davis et al., 2018)

Social interactions rely on core cognitive functions including attention, language, memory and executive functions (Beaudoin & Beauchamp, 2020). Core cognitive abilities and socio-cognitive abilities interact with each other throughout development (Beauchamp, 2017). For example, social skills like joint attention are crucial for language development (Derksen et al., 2018) while executive functions and language enhance socio-cognitive skills such as ToM (Matthews et al., 2018).

Leung et al. (2016) examined the association between specific executive functioning skills and SR (as measured by the SRS) in 70 children with autism and 71 without autism aged 6–15 years old. They reported difficulties in inhibition, shifting, and emotional control were associated with poorer SR in both groups. Initiation, working memory, planning, organisation, and monitoring difficulties, were linked to poorer SR in the children with autism only. A study by Torske et al. (2018) utilised the same measures as Leung et al. (2016) and reported a stronger relationship between executive functioning and SRS scores in autistic females compared to autistic males aged 6-18.

A study completed by Dai et al. (2019) investigated the associations between executive function and SRS scores in typically developing children, (n=413) aged 6-9 years. They also found sex specific associations between executive function and SRS scores but only in children with high (worse) SRS scores. Among males with higher scores, only set shifting predicted individual difficulties in scores, whereas all the components of executive functioning predicted variances of difficulties in females with higher SRS scores. These results indicate that specific executive functioning abilities and cognitive domains may be more closely associated with SRS scores depending on biological sex. However, further research is required with larger cohorts of typically developing children and adolescents. Despite the research which suggests and supports that there is an interactional relationship between core cognitive functions and social cognition (Beaudoin & Beauchamp, 2020), compared to executive functioning, less is known about the relationship between SR and other cognitive domains such as memory and language warranting exploratory research.

The current study used the Adolescent Brain Cognitive Development (ABCD) study data. The ABCD is a longitudinal study involving 11,875 children aged 9-10 years old at baseline. The aims of the present study were to describe SR in a large cohort of 9–12-year-olds. The study also aimed to determine if cognitive domains including executive functioning and social factors such as socio-economic status (SES) are associated with social responsiveness.

Research Questions

1. Are there age and sex differences in social responsiveness in children aged 9-12 years old and does age moderate the relationship between sex and social responsiveness?
2. To what extent are SES and performance on cognitive tests measuring abstract reasoning, executive functioning, cognitive flexibility, inhibitory control and attention, processing speed, episodic memory, working memory, verbal memory, receptive vocabulary, reading and decoding, associated with social responsiveness scores after accounting for age and sex?

Methods

Study Design

The data was taken from the ABCD study; a large longitudinal study of brain development and child health that is set to last 10 years. A total of 11,875 children were recruited aged 9-10 years at baseline between the 1st of October 2016 and 31st of October 2018. To create a diverse sample, participants were recruited from 21 research sites across the USA. The study includes a range of data types including performance-based measures, questionnaires, and neuroimaging procedures. Exclusionary diagnoses at the time of recruitment included a current diagnosis of schizophrenia, a moderate/severe autism diagnosis, intellectual disability, or alcohol/substance use disorder. A past diagnosis that had remitted was not exclusionary.

The study received approval from a centralised institutional review board from the University of California and San Diego. Each of the 21 study sites obtained local institutional review board approval. Informed written consent was obtained from parents, and children provided written assent. The study design and recruitment strategy are described fully by Garavan et al. (2018).

The University of Glasgow has institutional approval to use the ABCD data. To comply with data regulations, ABCD data was stored and analysed on University systems only. To gain approved access to the data, the principal researcher and her university supervisors were added to the 'list of approved researchers' and created a National Institute of Mental Health Data Archive (NDA) account (see Appendix N)

Participants

Only unrelated participants were included in this study. One member of each sibling group was randomly selected to remain in the analysis to retain as much of the sample as possible. The precedent of including unrelated participants only has been set by a previous study which has utilised SRS data from the ABCD study (Sharp et al., 2023). Statistical models were conducted with and without those with Autism Spectrum Disorder (ASD) and Attention Deficit Hyperactivity Disorder (ADHD). Please see the data analysis section below for further information about how participant exclusions were handled in the analyses. Participant's SRS data was taken from year 1, and information about exclusionary diagnoses was taken from baseline and the year 1 follow-up visit. Data for all other measures was taken from the baseline visit.

Materials and Measures

Demographic and Diagnostic Information

Child sex, age, whether the child was born in the USA and total household income (as an indicator of SES), was extracted from the demographics survey which comprised items mostly from the PhenX toolkit (Stover et al., 2010) answered by the child's main caregiver at baseline. In accordance with previous research using ABCD data (Cullen et al., 2023), born in the USA was used as a proxy for likely English-language fluency, which was added as a covariate to the analyses involving the cognitive scores.

Information about ASD diagnoses was taken from the 'ASD Screener' at baseline; although 'moderate/severe autism' was an exclusion criterion, a small number of eligible participants did have a diagnosis of ASD recorded on this screener. Information about ADHD diagnoses was taken from the baseline screener (in which it was grouped alongside other mental health conditions) and the parent/caregiver-reported Schedule for Affective Disorders and Schizophrenia for School-Age Children–Computerized version (K-SADS-COMP) present and lifetime version (Kaufman et al., 1997) administered at both baseline and year 1. The screener results pertaining to ADHD and other conditions are reported here for information only; the ADHD analysis exclusions were based only on the K-SADS data.

The Social Responsiveness Scale (SRS) Abridged Version **(Parent/Carer Report)**

Social responsiveness was assessed using data from an abridged 11-item version of the SRS to reduce participant burden (Reiersen et al., 2008). This was administered to parents/carers at the one-year follow-up only (not at baseline). The original version of the SRS is used to assess the severity of social difficulties in children with and without ASD aged 4-18-years (Constantino & Gruber, 2005). Parents/caregivers were asked to rate statements on a four-point Likert scale: 0 (not true); 1 (sometimes true); 2 (often true); and 3 (almost always true). Most of the SRS items were related to deficits in reciprocal social behaviour, but there were also three items related to stereotyped/repetitive behaviours (SRS items 24, 29, and 39) and an item related to communication impairment (SRS item 35). Higher scores indicate poorer social responsiveness.

Cognitive Assessments

Episodic memory, executive function, attention, working memory, processing speed, language abilities and overall cognitive function were measured using scores from the National Institutes of Health Toolbox (NIH-TB) cognitive battery (McDonald, 2014). The NIH-TB cognitive battery is a brief assessment tool designed for large epidemiological and longitudinal studies and is validated for use with participants aged 3–85-years. The NIH-TB cognitive battery was administered at baseline in person via an iPad and is comprised of seven tasks. This has been validated against gold-standard neuropsychological assessments and demonstrated good test-retest reliability and validity in children and adolescents (Weintraub et al., 2013). These scores were age-corrected centrally by ABCD and are reported as standard scores with a mean of 100 and SD of 15 (higher = better).

The Rey Auditory Verbal Learning Test (RAVLT) Immediate Recall and Delayed Recall subtests were also administered in person at baseline (Strauss et al.,2006). Raw scores were converted into age-corrected z-scores based on the analysis sample distribution (mean 0, SD 1; higher = better). The Matrix Reasoning task from the Wechsler Intelligence Scale for Children (WISC-V; Weschler, 2014) was also administered at baseline, with scores reported as age-corrected scaled scores with a mean of 10 and SD of 3 (higher = better).

Table 1 below outlines the cognitive domains measured by each task.

Table 1. Name of task and Associated Domain

Task Name	**Cognitive Domain**
NIH Picture Vocabulary	Receptive Vocabulary
NIH Flanker	Attention and inhibitory control
NIH List sorting	Working memory
NIH Card sorting	Set Shifting
NIH Pattern Comparison	Processing speed
NIH Picture Sequence	Episodic Memory
NIH Oral Reading	Verbal Expression
RAVLT	Verbal Learning and Memory
WISC Matrix Reasoning	Visual Information Processing and Abstract Reasoning

Abbreviations: NIH: National Institutes of Health

Covariates

Child Behaviour Checklist (CBCL)

The CBCL is a 113- item, parent report measure designed to assess behavioural and emotional problems in 6–18-year-olds. Higher scores indicate increased behavioural and emotional problems. Mean test-retest reliabilities have been reported to range from 0.95 to 1.00, and internal consistency has ranged from 0.78 to 0.97 (Achenbach, 2009).

Data Analysis

The data analysis plan was developed a priori based on the research questions and considering how ABCD data analyses has been conducted by other researchers using the same variables, for example, Sharp et al. (2023). Descriptive statistics were used to describe sample demographics and participants' scores on the measures. A two-way between groups ANOVA was used to determine age and sex differences in SRS scores. Post-hoc tests were not required. A series of multiple regression models were completed to determine to what extent socioeconomic and cognitive measures are associated with SR scores in this cohort, controlling for age, sex and born in USA. A complete-case analysis was conducted, and missing data was not imputed. Unless otherwise stated below, the significance level was set as $p < .05$. A separate ANOVA analysis and regression models were completed with and without those with ASD and/or ADHD because the SRS is used to screen for ASD in clinical contexts and these diagnoses may have impacted scores on some of the clinical items. Furthermore, core cognitive domains such as executive functioning are known to be implicated in those with ADHD, and therefore may have skewed the results.

Scores from the CBCL administered at baseline were added to secondary analyses as a covariate to adjust for the influence of problem behaviours on SRS scores (Hus et al., 2013). This precedent was set by previous research that included the T-scores for internalising and externalising problems in their analyses of the ABCD SRS scores (Sharp et al., 2023). Each iteration of the regression models was completed again with the inclusion of the CBCL.

In addition to the main models, sensitivity analyses were completed to check for the potential influence of unrepresentativeness in the sample. There is evidence that the ABCD sample slightly over-represents two-parent families and higher-income households and under-represents some ethnic minority groups (Heeringa & Berglund, 2020). To reduce bias, ABCD created population weights for each sample member, so that researchers had the option to adjust their analyses to be more representative of the population. These weights were based on a comparison to the American Community Survey run by the US Census Bureau. There is no consensus as to whether researchers must use these weights, but for the purpose of comparison, it was decided that this study would repeat all analyses using the weights to see if this changed the results. This was done using the Complex Samples functions in SPSS software, taking account of the sample weights and study sites from which participants were recruited. These results are referred to below as population-weighted analyses.

Statistical Power

Given that the likely sample size was already known in advance of this study, a post-hoc power calculation was completed to determine statistical power for the primary research question. Taking the significance criterion (alpha) to be 0.05 and effect size to be small ($f = 0.1$), with four age groups and two sex groups, the calculation indicated that a sample of 7900 would provide power of 1.0 for a two-way between-groups ANOVA, including testing for main effects and interactions. The very large sample size also provided high power for the secondary research question using multiple regression.

Results

Sample Characteristics

Sample characteristics at baseline are outlined in Table 2. Sample characteristics for year 1 follow-up are outlined in Table 3. Descriptive statistics for scores on the SRS are shown in Table 4. Descriptive statistics for the psychological variables are shown in Appendix C.

Table 2.

Sample Characteristics at Baseline.

Variable: Number of completed data (Missing data)	Number of Participants (% of respondents)
Sex: N=9804	
Male	5143 (52.5%)
Female	4661 (47.5%)
Age: N=9803 (1)	
8 years	118 (1.2%)
9 years	5079 (51.8%)
10 years	4503 (45.9%)
11 years	103 (1.1%)
Born in the US: N=9,790 (14)	
Yes	9474 (96.6%)
No	316 (3.2%)
ASD Diagnosis Screener: N=9,777 (27)	
Yes	161 (1.6%)
No	9616 (98.1%)
ADHD (KSADS): N=9,804	
Yes	7 (0.1%)
No	9797 (99.9%)
ADHD and Other Mental Health Difficulty (Screener): N=9,779 (25)	
Yes	1452 (14.8%)
No	8327 (84.9%)
Household Income: N=8,948 (856)	
Less than $5,000	357 (3.6%)
$5,000- $11,999	365 (3.7%)
$12,000- $15,999	239 (2.4%)

$16,000- $24,999	436 (4.4%)
$25,000- $34,999	566 (5.8%)
$35,000- $49,999	770 (7.9%)
$50,000- $74,999	1216 (12.4%)
$75,000- $99,999	1304 (13.3%)
$100,000- $199,999	2683 (27.4%)
$200,000 and above	1012 (10.3%)

Abbreviations: ADHD, Attention Deficit Hyperactivity Disorder; ASD, Autism Spectrum Disorder; KSADS, Kiddie Schedule for Affective Disorders and Schizophrenia.

Table 3.

Sample Characteristics at Year 1 Follow-up. †

Variable: Number of completed data (Missing data) ‡	**Number of Participants (% of respondents)**
Age: N=9,234 (570)	
8 years	0 (0%)
9 years	378 (3.9%)
10 years	4504 (45.9%)
11 years	4033 (41.1%)
12 years	319 (3.3%)
ADHD KSADS: N=9234 (570)	
Yes	10 (0.1%)
No	9224 (94.1%)

Abbreviations: ADHD, Attention Deficit Hyperactivity Disorder; KSADS, Kiddie Schedule for Affective Disorders and Schizophrenia.

†: Screener questions were not administered at year 1.

‡: Missing data at Year 1 is because some participants did not attend Year 1 assessment.

Table 4.

Social Responsiveness Scores at year 1†

Total		
N=9222	**14.44 (4.13)**	
Age	**Male**	**Female**

9 years	14.49 (3.71)	13.74 (3.48)
10 years	14.91 (4.54)	13.87 (3.46)
11 years	14.92 (4.54)	14.03 (3.66)
12 years	14.84 (5.05)	13.95 (3.57)

†: All data are mean (standard deviation)

Are there age and sex differences in social responsiveness scores?

A two-way between-groups analysis of variance was conducted to explore the impact of age and sex on SRS scores at Year 1 follow-up. Participants were divided into four groups depending on their age at follow-up (Group 1: 9 years, Group 2: 10 years, Group 3: 11 years, and Group 4: 12 years). The significance level was set to 0.01 because Levene's Test of Equality of Error Variances was violated. The interaction effect between sex and age group was not statistically significant, F (3, 9213) =.337, p=.798. There was not a statistically significant main effect for age, F (3, 9213) =1.03, p=.374. There was a statistically significant main effect for sex, F (1, 9213) =30.2, p=.001; however, the effect size was small (η^2_p=.003). Table 3 shows the mean SRS score per age group by sex, which indicates that the significant result was driven by slightly higher SRS scores in males.

When this analysis was repeated after excluding participants with ASD and/or ADHD, the results were similar (see Appendix D1). This analysis was repeated as part of the sensitivity analyses using population weighting and the results were similar (see Appendices D2, D3).

To what extent are socioeconomic and cognitive measures associated with social responsiveness scores?

A series of multiple regression models were completed with and without participants with ASD and/or ADHD. Separate regression models were also completed with males and females only, and further models were completed including CBCL scores as additional covariates. Sensitivity analyses were completed for each iteration of the regression models using population- weighted analyses.

There were indications of multicollinearity among the cognitive scores in all the regression models, with variance inflation factor (VIF) values being higher than the recommended cut-off of 10 (Pallant, 2020). There was a high VIF of 1204 for the NIH Toolbox total cognitive score, derived from a composite of the separate NIH Toolbox domain scores. This score was also found to lead to implausibly high coefficients, in the opposite direction to what was expected. This composite score was therefore removed from all models. However, it was decided to retain all the separate cognitive scores in the models to differentiate the associations with SRS scores across different cognitive domains. The VIF scores were checked for subsequent models without NIH total and were found to be well below the threshold of 10 (highest VIF = 3, for both RAVLT scores). All the other regression model assumptions were met.

The first regression model assessed the association of age, sex, born in the US (as a proxy for English-language fluency), household income (as an indication of socio-economic status), scores on the seven NIH toolbox tasks, matrix reasoning and RAVLT verbal memory, with the dependent variable SR score in all unrelated participants. This was a statistically significant model *F* (14, 8015=27.92, p<.001). The adjusted R^2 indicated that 4.5 percent of the variance in SR was explained by variances in these predictor variables. The results for each predictor are shown in Table 5. The analysis suggested that male sex (η^2_p=.013) had the strongest association with SRS scores. Male sex, being born in the US, lower household income, higher scores on the NIH picture vocabulary and lower scores on the NIH flanker, list sorting, oral reading and RAVLT immediate recall were shown to be statistically significant predictors of higher SRS scores.

Table 5.

Multiple Regression Analysis for Social Responsiveness Scores

Predictor Variable	*B*(unstandardised)	Confidence Interval (95%)	t	Sig.	Partial eta-squared
Intercept	16.715	15.033, 18.396	19.485	**<.001**	.045
Sex (male)	.934	.758, 1.111	10.364	**<.001**	.013
Born in US (no)	-.547	-1.057, -.037	-2.103	**.036**	.001
NIH Picture Vocabulary	.015	.009, .021	4.665	**<.001**	.003
NIH Flanker	-.017	-.025 , -.010	-4.601	**<.001**	.003
NIH List Sorting	-.008	-.015, -.001	-2.352	**.019**	.001
NIH Card Sorting	-.003	-.010, .004	-.881	.378	.000
NIH Pattern Comparison	-.002	-.007, .002	-.995	.320	.000
NIH Picture Sequence	-.001	-.007, .005	-.216	.829	.000
NIH Oral Reading	-.006	-.011, .000	-2.031	**.042**	.001
RAVLT IM	-.282	-.443, -.120	-3.423	**<.001**	.001
RAVLT DL	.007	-.156, .170	.079	.937	.000
WISC Matrix Reasoning	.018	-.015, .052	1.071	.284	.000
Household Income	-.202	-.243, -.162	-9.827	**<.001**	.012
Age	.051	-.091, .193	.704	.481	.000

Adjusted R^2 =4.5 percent; (*F* (14, 8015=27.92, *p*<.001). Sample is unrelated participants including those with ASD/ADHD diagnoses.

Abbreviations: NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

When this analysis was repeated after excluding participants with ASD and/or ADHD, being born in the US was no longer a statistically significant predictor (see Appendix E1). When this analysis was repeated using population weighting, the model accounted for slightly more variance at 5.5 percent. The same variables remained statistically significant (including born in the US), except for NIH list sorting when including all participants (see Appendix E2). On the population-weighted version excluding participants with ASD and/or ADHD, NIH list sorting was a statistically significant predictor, but RAVLT immediate recall was not (see Appendix E3).

Sex-specific Associations

To determine if there are sex-specific associations for predictors, separate sex-specific regression models were completed, each including the predictor variables age, born in the US (as a proxy for English-language fluency), household income (as an indication of SES), scores on the seven NIH toolbox tasks, matrix reasoning and RAVLT verbal memory, with the dependent variable SR score.

The male only model was a statistically significant model (F (13, 4201=12.59, p<.001). The adjusted R^2 indicated that 3.5 percent of the variance in SR can be explained by variances in the predictor variables. The results for each predictor are shown in Table 6. The analysis suggested that household income (η^2_p=.014) had the strongest association with SRS scores. Lower household income, being born in the US, lower scores on the NIH Flanker, NIH list sorting and RAVLT immediate recall tasks and higher scores on the NIH picture vocabulary and WISC matrix reasoning tasks were shown to be statistically significant predictors of higher SRS scores.

Table 6.

Multiple Regression Analysis Summary Predicting Social Responsiveness Scores in Males

Predictor Variable	*B*(unstandardised)	Confidence Interval (95%)	t	Sig.	Partial eta-squared
Intercept	18.997	16.404, 21.550	14.462	**<.001**	.047
Born in US (no)	-1.072	-1.880, -0.264	-2.601	**.009**	.002
NIH Picture Vocabulary	.018	.009, .028	3.663	**<.001**	.003
NIH Flanker	-.022	-.033, -.011	-3.997	**<.001**	.004
NIH List Sorting	-.015	-.026, -.004	-2.711	**.007**	.002
NIH Card Sorting	-.005	-.015, .006	-.904	.366	.000
NIH Pattern Comparison	.001	-.006, .007	.214	.831	.000
NIH Picture Sequence	-.001	-.011, .008	-.251	.802	.000
NIH Oral Reading	-.003	-.012, .005	-.818	.413	.000
RAVLT IM	-.338	-.585, -.090	-2.677	**.007**	.002
RAVLT DL	-.005	-.253, .244	-.037	.971	.000
WISC Matrix Reasoning	.063	.012, .114	2.418	**.016**	.001
Household Income	-.238	-.299, -.177	-7.657	**<.001**	.014
Age	-.052	-.268, .164	-.474	.635	.000

Adjusted R^2 *=3.5 percent; (13, 4201=12.59, p<.001).* Sample is unrelated male participants including those with ASD/ADHD diagnoses.

Abbreviations: NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

When this analysis was repeated after excluding male participants with ASD and/or ADHD, the results were similar (see Appendix F1). When these results were repeated using population weighting, the model accounted for slightly more variance at 4.6 percent, but picture vocabulary and list sorting were no longer statistically significant predictors (see

Appendix F2). When the population-weighted analysis was completed again excluding participants with ASD and/or ADHD, list sorting was a statistically significant predictor, but RAVLT was not (see appendix F3).

The female only regression model was a statistically significant model (*F* (13, 3801=9.84, *p*<.001). The adjusted R^2 indicated that 2.9 percent of the variance in SR can be explained by variances in the predictor variables. The results for each predictor are shown in Table 7.. The analysis suggested that household income (η^2_p=.010) had the strongest association with SRS scores. Lower household income, lower scores on the NIH flanker, pattern comparison, oral reading and RAVLT immediate recall tasks, and higher scores on the NIH picture vocabulary task, were shown to be statistically significant predictors of higher SRS scores.

Table 7.

Multiple Regression Analysis Summary Predicting Social Responsiveness Scores in Females

Predictor Variable	*B*(unstandardised)	Confidence Interval (95%)	t	Sig.	Partial eta-squared
Intercept	15.222	13.113, 17.331	14.149	**<.001**	.050
Born in US (no)	-.038	-.654, .578	-.121	.903	.000
NIH Picture Vocabulary	.012	.004, .019	2.975	**.003**	.002
NIH Flanker	-.011	-.021, -.001	-2.224	**.026**	.001
NIH List Sorting	-.002	-.011, .007	-.483	.629	.000
NIH Card Sorting	-.001	-.009, .008	-.167	.867	.000
NIH Pattern Comparison	-.006	-.012, .000	-2.108	**.035**	.001
NIH Picture Sequence	.000	-.008, .008	-.026	.979	.000
NIH Oral Reading	-.008	-.015, -.001	-2.274	**.023**	.001
RAVLT IM	-.224	-.425, -.022	-2.176	**.030**	.001
RAVLT DL	.025	-.180, .229	.235	.814	.000
WISC Matrix Reasoning	-.030	-.072, .012	-1.385	.166	.001
Household Income	-.164	-.215, -.113	-6.266	**<.001**	.010

Age	.165	-.013,	.343	1.814	.070	.001

Adjusted R^2 =2.9 percent; (F (13, 3801=9.84, p<.001). Sample is unrelated female participants including those with ASD/ADHD diagnoses.

Abbreviations: NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

When this analysis was repeated after excluding female participants with ASD and/or ADHD, the results were similar (see Appendix G1), although NIH flanker scores were not a statistically significant predictor. The population-weighted model accounted for slightly more variance at 4.2 percent. In the population-weighted analysis, higher scores on NIH picture vocabulary, lower scores on the NIH oral reading task and lower household income were statistically significant predictors (see Appendix G2). This was the same for the population-weighted analysis excluding participants with ASD and/or ADHD (see appendix G3).

Impact of Co-occurring Psychopathology

To assess the impact of co-occurring psychopathology on SRS scores, additional regression models were completed with the inclusion of T-scores of externalising and internalising symptoms extracted from the CBCL.

The first regression model included the predictors from the previous models with the addition of CBCL externalising and internalising scores, with the dependent variable SR score. This was a statistically significant model (*F* (16, 8012=155.60, *p*<.001). The adjusted R^2 indicated that 23.6 percent of the variance in SR was explained by variances in these predictor variables. The results for each predictor are shown in Appendix H1. The analysis suggested that CBCL internalising scores (η^2_p=.073) had the strongest association with SRS scores. Male sex, lower household income, higher scores on the NIH picture vocabulary, higher CBCL externalising and internalising scores and lower scores on the NIH flanker, list sorting, oral reading and RAVLT immediate recall were shown to be statistically significant predictors of higher SRS scores.

When this analysis was repeated after excluding participants with ASD and/or ADHD, NIH list sorting was no longer a statistically significant predictor (See Appendix H2). When these results were repeated using population-weighted analyses, the results were similar (see Appendices H3, H4)

The separate regression models for male and females were repeated with CBCL scores. The male regression model including CBCL scores was a statistically significant model (F (15, 4198=88.83, p<.001). The adjusted R^2 indicated that 23.8 percent of the variance in SR was explained by variances in these predictor variables. The results for each predictor are shown in Appendix I1. The analysis suggested that CBCL internalising scores (η^2_p=.086) had the strongest association with male SRS scores. Being born in the US, lower household income, higher scores on the NIH picture vocabulary, higher CBCL externalising and internalising scores and lower scores on the NIH flanker, list sorting, RVALT immediate recall and matrix reasoning tasks were shown to be statistically significant predictors of higher SRS scores.

When this analysis was repeated after excluding male participants with ASD and/or ADHD, NIH list sorting and WISC matrix reasoning were no longer statistically significant predictors. NIH oral reading became a statistically significant predictor (see Appendix I2.) When the analyses were repeated using population weighting, born in the US, lower household income, lower NIH flanker and RAVLT immediate recall scores and higher scores on both CBCL scales and matrix reasoning were statistically significant predictors (see Appendix G3). When the population-weighted analysis was repeated for excluding people with ASD and/or ADHD, RAVLT immediate recall and matrix reasoning scores were no longer statistically significant (see Appendix I4)

The female regression model including CBCL scores was a statistically significant model (F (15, 3799=69.89, p<.001). The adjusted R^2 indicated that 21.3 percent of the variance in SR was explained by variances in these predictor variables. The results for each predictor are shown in Appendix J1. The analysis suggested that CBCL internalising scores (η^2_p=.061) had the strongest association with female SRS scores. Lower household income, lower scores on the NIH oral reading and RAVLT immediate recall tasks, higher CBCL externalising and internalising scores and age were shown to be statistically significant predictors of higher SRS scores.

When this analysis was repeated after excluding female participants with ASD and/or ADHD, NIH pattern comparison became a statistically significant predictor (see Appendix J2). When the analyses were repeated using population weighting, lower household income, lower oral reading scores and higher scores on both CBCL scales were statistically significant predictors (see Appendices J3, J4).

Post-hoc Analyses for Picture Vocabulary and Matrix Reasoning

Considering the counterintuitive result which showed that the association between picture vocabulary and SRS was in the opposite direction from the other cognitive scores, an additional sensitivity analysis was conducted which included picture vocabulary and demographic predictors but omitted the other cognitive scores (in the full unrelated sample). This analysis demonstrated that the positive association between better picture vocabulary abilities and higher (worse) SRS was absent in this model and was therefore only evident after adjusting for other cognitive domains (see Appendix K1).

A sensitivity analysis was also completed for matrix reasoning in the male only regression model, because this had also shown an unexpected positive association with higher SRS. The analysis included matrix reasoning and demographic predictors but omitted the other cognitive scores (in the male unrelated sample). This analysis demonstrated that the positive association between better matrix reasoning abilities and higher (worse) SRS was also only evident after adjusting for other cognitive domains (see Appendix K2)

Discussion

The primary analysis showed there was not a significant increase in social responsiveness (SR) ability from ages 9-12 years. There was a statistically significant female advantage across the age range which was driven by slightly higher SRS scores in males. However, in light of the small effect size, the significant sex difference found in the present study should be interpreted with caution, as the large sample size may have increased the statistical power, potentially leading to the detection of minor effects that do not hold practical or clinical significance.

Previous research has reported females tend to score lower on the SRS compared to males, but not significantly so in typically developing children aged 11-18, 6-12 and 4-17 years in samples ranging from 49-438 participants (Backer Van Ommeren et al., 2017; Sedgewick et al., 2019; Wallace et al., 2017). In a recent validation study of a shortened 16-item SRS involving 7,030 participants, a statistically significant sex difference was detected (with higher scores among males) (Kaat et al., 2023). Like the present study, the reported effect size was also small indicating statistically significant sex differences may only emerge in studies with large samples.

There was not a statistically significant effect for age, which suggests between the ages of 9-12, the SRS is possibly less sensitive to detecting significant increases or decreases in SR in this cohort. However, there was significantly less SRS data for the 9- and 12-year-olds which may have affected the representation of the youngest and oldest participants and hence the ability to detect significant differences in scores between the age groups. There was not a statistically significant interaction effect between age and sex, which indicates the slight female advantage is present regardless of age, consistent with results from a recent systematic review (Wood-Downie et al., 2021).

Predictors of Social Responsiveness Scores

The socioeconomic and cognitive predictor variables accounted for little variance in SRS scores; the statistical significance of predictor variables varied across different regression iterations. However, further inspection of the models demonstrated the changes in statistical significance did not coincide with substantive changes in effect size or confidence intervals. Thus, changes in statistical significance likely reflect the models' high power and sensitivity around the significance threshold rather than any meaningful impact of excluding participants with ASD and/or ADHD for example. Therefore, the discussion will focus on the most consistent predictors and results of interest across all the models.

Demographic Factors

As expected, male sex and lower household income were statistically significant predictors of higher (poorer) SRS scores in most of the raw and population-weighted regression models. Being born in the US was also associated with higher SRS scores in most of the raw and population-weighted models including all participants and the male specific model. This may be reflective of 'westernised' biases in the observation and evaluation of pre-adolescent male social behaviour as described by Whitlock et al., (2020) and an increase in the pathologisation of child and adolescent behaviour in general (Waite-Jones& Rodriguez, 2022). However, no initial hypothesis was generated in relation to this variable (as it was included primarily as a potential confounder for cognitive performance), therefore, further research with a theoretical basis would be required to fully understand this result.

Cognitive Scores

The cognitive scores accounted for relatively little variance in the regression models. Some cognitive domains were statistically significant predictors of SR scores; however, these had very small effect sizes indicating the effect they have on social responsiveness is minimal. Consistent with previous research, poorer oral reading ability and attentional and inhibitory control were statistically significantly associated with poorer SR in the raw and population-weighted models (Beaudoin & Beauchamp; Leung et al., 2016)

It was anticipated that cognitive scores would account for more variance given the numerous studies that report a relationship between social and core cognitive abilities (Beaudoin and Beauchamp, 2020). However, other studies may exist that have found a non-significant relationship but were not published, which highlights the wider issue of publication bias's influence on research hypotheses and scientific understanding.

Higher scores on the picture vocabulary task (which measures receptive vocabulary performance) were associated with poorer SR. This unexpected, counterintuitive result was replicated in the population-weighted models. This contrasts with previous research that found higher picture vocabulary scores were associated with lower (better) scores on the SRS in pre-school children (Cheung et al., 2022). However, when picture vocabulary was analysed alone without other cognitive measures, this counterintuitive association disappeared. Therefore, the link between better picture vocabulary performance and worse SR was only apparent in conjunction with the other cognitive scores. There has been no past research which has reported a similar finding, and it cannot be attributed to multicollinearity as VIFs were acceptably low in these models. The association between picture vocabulary performance and other cognitive measures suggests that SR and its links with core cognitive abilities involves complex interactions.

Sex-specific Associations

The separate female and male regression models accounted for very similar amounts of variance. There were some differences in the predictor variables that were significant for males versus females, however, these had small effect sizes.

Attentional and inhibitory control (as measured by the NIH Flanker task) was only a statistically significant predictor in the male models. Sex-specific associations between executive functioning domains and SR abilities as measured by the SRS have been reported in cohorts with ASD (Torske et al., 2018) and in subgroups with higher (worse) SRS scores in typically developing samples (Dai et al., 2019). Torske et al. (2018) reported stronger associations between SRS scores and executive functioning in females compared to males. Dai et al. (2019) reported among males with higher SRS scores, only set shifting predicted individual difficulties in social awareness, whereas all the components of executive functioning predicted variances of difficulties in social cognition and social communication in females with higher SRS scores. However, they found no significant associations between executive functioning and SR in males and females those with lower (better) SRS scores.

These findings differ with the present study, as neither of the tasks that measure executive functioning (NIH Dimensional sort task, NIH Flanker task) were significant predictors of female SRS scores, and only the attentional and inhibitory control domains of executive functioning were significant predictors of male SRS scores. The difference in results could be a result of methodological differences or may reflect the developmental instability of executive functioning. A three-year longitudinal study examined the relationship between social function (as measured by the SRS) and executive function across adolescence in a sample of 9–18-year-olds at baseline. The study reported executive functioning ability only predicted social cognition at Year 1, suggesting the influence of executive functioning on social cognition changes overtime (Ben-Asher et al., 2023).

Higher (better) scores on the WISC Matrix Reasoning task (which measures visual information processing and abstract reasoning), was a statistically significant predictor of poorer SR among males only. This counterintuitive result was replicated in the population-weighted analyses. However, when matrix reasoning was analysed alone without other cognitive measures, this counterintuitive association disappeared. Therefore, the link between better matrix reasoning performance and worse social abilities was only apparent in conjunction with the other cognitive scores. This finding cannot be attributed to multicollinearity as VIFs were acceptably low in these models. This finding is difficult to explain, and previous research has not reported a similar pattern. This result could be considered in the context of broader research on sex differences in cognitive abilities. Meta-analytic studies indicate that females have a cross-cultural advantage in reading, writing, language, and verbal memory, whereas males are significantly over-represented in studies examining superior mathematical and spatial reasoning abilities (see Miller & Halpern, 2014, for a review). For the counterintuitive association between better matrix reasoning and worse SR when measured with other cognitive domains to be conclusively confirmed as genuine, this finding would need to be replicated in a different representative sample.

Lastly, poorer oral reading (which measures verbal expression abilities) was statistically significantly associated with poorer SR among females only. Verbal communication is likely important for socialising in both sexes, but females tend to utilise more language-related strategies in social situations (Gutierrez et al., 2020). Therefore, the impact of poorer oral reading skills on SR may be more pronounced in females because their social interactions rely heavily on verbal communication.

The sex specific associations between the cognitive tests and SRS scores do not add to the understanding of the female advantage. Research on androgen expression in-utero offers a compelling biological explanation for the advantage, particularly because these differences occur before a child is exposed environmental influences. Androgen exposure in-utero has been linked to more male-stereotypical cognition and behaviour (Baron-Cohen et al., 2015) and poorer interpersonal skills and empathy (Knickmeyer & Baron-Cohen, 2006).

Furthermore, a recent meta-analysis of 20 independent studies suggests prenatal androgen exposure masculinises and defeminises play behaviour (Kung et al., 2024).

It is also important to consider theories such as Social Role Theory which suggest differences in social behaviour between males and females are a result of the environmental shaping of gender roles (Eagly & Wood, 2012), and research has suggested differences between males and females exist in social cognitive research because of adherence to social norms rather than innate female advantage (Gutierrez et al., 2020). The debate is even more complex given the increasing awareness of the differences between sex and gender. To understand the female advantage, it is important for researchers to consider the difference between sex and gender. Generally, sex refers to a set of biological attributes, while gender refers to the socially constructed roles, behaviours, and identities of females, males, and gender diverse people (Heidari et al., 2016). This differentiation is important in the ongoing debate as to whether the female advantage is fixed due to biological differences or whether social and environmental factors also influence social abilities.

Influence of Problem Behaviours

The regression models including the CBCL scores accounted for much more variance in SRS scores, which was expected based on previous research that found SRS scores are highly influenced by problem behaviours (Hus et al., 2013). The statistically significant predictors in the models without the CBCL remained statistically significant. Several of the associations weakened, indicating the apparent links between demographic factors and SR was at least partly accounted for by the co-occurrence of behavioural problems. However, these were minimal changes in effect size.

Parent-reported measures of social difficulties and problem behaviours are often highly correlated due to the subjective nature of parental observations. Parents perceptions can be influenced by their child’s overall difficulties, leading to overlapping ratings across different domains (Denham et al., 2020). For example, a child exhibiting problem behaviours, such as aggression or defiance, may struggle with social interactions which can result inflated correlations between constructs that are supposed to measure the different domains (Achenbach, 2006).

In contrast, performance-based measures such as the cognitive assessment used in this study are designed to assess specific abilities. These objective assessments tend to show weaker correlations with parent-reported measures because they are less susceptible to the broad generalisations parents may make in naturalistic observations (Denham et al., 2021). This may help explain why the cognitive scores did not account for as much variance in SRS scores as expected.

Study Limitations and Strengths

The main strengths of this study include the large sample size and use of population-weighted analysis to determine the generalisability of the findings. The inclusion of the CBCL as a measure of problem behaviour in additional regression models allowed for a more nuanced interpretation of the results which may have been confounded by co-occurring behavioural difficulties.

A strength of the study is the use of the SRS, which is recommended by the RDoC framework to assess social abilities in typically developing young people, but a limitation is that the version of the SRS that was used was abridged. The use of abridged psychometric tools is a common compromise in large research studies to reduce participant burden. However, the abridged SRS has not been validated and therefore, may have reduced sensitivity or specificity compared to the full version. Furthermore, the SRS scores and cognitive scores were taken from two different time points (1 year apart). The participants' abilities in the cognitive domains are likely to have developed during this period. Therefore, the association between SRS and cognitive scores are not as closely linked as they might be if they both were taken at the same time point. On the other hand, the fact that the cognitive assessments pre-dated the SRS measure may facilitate interpretation of the directionality of any associations in terms of potential causality. Lastly, the use of ‘born in the USA’ was an imperfect proxy for English language fluency but there was no direct measure of English language fluency in this data set.

Future Research

The variables analysed in the present study accounted for very little variance in the regression models; therefore, future studies should focus on other factors such as in-utero androgen exposure and socio-cultural influences to broaden our understanding of the potential mechanisms of social development in pre-adolescents. This study did find differences in the patterns of associations with variables for males and for females, but further research is required to develop a comprehensive model that explains the female advantage.

The two counterintuitive associations for picture vocabulary and matrix reasoning cannot be explained by previous research. These associations had very small effect sizes which means they hold minimal clinical or practical importance; however, future research could attempt to replicate this finding and the reasons behind it based on a theoretical framework. Finally, as the ABCD cohort is an ongoing, longitudinal study, it provides the ideal sample to continue examining these trends as further data is released.

Conclusions

This study describes SR in large cohort of 9–12-year-olds and confirms previous findings in relation to the female advantage. The predictor variables included in the study provided little insight into the factors that predict SR in 9–12-year-olds or the mechanisms behind the female advantage. Further research is required to understand the factors that promote social development and the drivers behind the female advantage.

There were no sources of funding for this research.

References

Achenbach, T. M. (2006). As others see us: Clinical and research implications of cross-informant correlations for psychopathology. *Current directions in psychological science*, *15*(2), 94-98.

Achenbach, T. M. (2009). *The Achenbach system of empirically based assessment (ASEBA): Development, findings, theory, and applications*. University of Vermont, Research Center for Children, Youth, & Families.

Backer van Ommeren, T., Koot, H. M., Scheeren, A. M., & Begeer, S. (2017). Sex differences in the reciprocal behaviour of children with autism. *Autism*, *21*(6), 795-803.

Baron-Cohen, S., Auyeung, B., Nørgaard-Pedersen, B., Hougaard, D. M., Abdallah, M. W., Melgaard, L., ... & Lombardo, M. V. (2015). Elevated fetal steroidogenic activity in autism. *Molecular psychiatry*, *20*(3), 369-376.

Baron-Cohen, S., Radecki, M. A., Greenberg, D. M., Warrier, V., Holt, R. J., & Allison, C. (2022). Sex differences in theory of mind: The on-average female advantage on the Reading the Mind in the Eyes Test. Developmental medicine and child neurology, 64(12), 1440-1441.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The “Reading the Mind in the Eyes” Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. The Journal of Child Psychology and Psychiatry and Allied Disciplines, 42(2), 241-251.

Beauchamp, M. H. (2017). Neuropsychology’s social landscape: Common ground with social neuroscience. *Neuropsychology*, *31*(8), 981.

Beaudoin, C., & Beauchamp, M. H. (2020). Social cognition. In Handbook of clinical neurology (Vol. 173, pp. 255-264). Elsevier.

Ben-Asher, E., Porter, B. M., Roe, M. A., Mitchell, M. E., & Church, J. A. (2023). Bidirectional longitudinal relations between executive function and social function across adolescence. *Developmental psychology*, *59*(9), 1587.

Cheung, R. W., Hartley, C., & Monaghan, P. (2022). Receptive and expressive language ability differentially support symbolic understanding over time: Picture comprehension in late talking and typically developing children. *Journal of Experimental Child Psychology*, *214*, 105305.

Constantino, J. N. (2021). Social responsiveness scale. In Encyclopedia of autism spectrum disorders (pp. 4457-4467). Cham: Springer International Publishing.

Constantino, J. N., & Gruber, C. P. (2012). *Social responsiveness scale second edition (SRS-2): Manual*. Western Psychological Services (WPS).

Constantino, J. N., Davis, S. A., Todd, R. D., Schindler, M. K., Gross, M. M., Brophy, S. L., ... & Reich, W. (2003). Validation of a brief quantitative measure of autistic traits: comparison of the social responsiveness scale with the autism diagnostic interview-revised. *Journal of autism and developmental disorders, 33*, 427-433.

Cullen, B., Gameroff, M. J., Ward, J., Bailey, M. E., Lyall, D. M., Lyall, L. M., ... & Weissman, M. M. (2023). Cognitive function in people with familial risk of depression. JAMA psychiatry, 80(6), 610-620.

Dai, M., Lin, L., Liang, J., Wang, Z., & Jing, J. (2019). Gender difference in the association between executive function and autistic traits in typically developing children. *Journal of autism and developmental disorders*, *49*, 1182-1192.

Davis, A. N., Carlo, G., Streit, C., & Crockett, L. J. (2018). Considering economic stress and empathic traits in predicting prosocial behaviors among US Latino adolescents. *Social Development*, *27*(1), 58-72

Denham, S. A., Ferrier, D. E., Howarth, G. Z., Herndon, K. J., & Bassett, H. H. (2020). Key considerations in assessing young children's emotional competence. In Social and emotional learning (pp. 27-45). Routledge.

Derksen, D. G., Hunsche, M. C., Giroux, M. E., Connolly, D. A., & Bernstein, D. M. (2018). A systematic review of theory of mind's precursors and functions. *Zeitschrift für Psychologie*.

Dorris, L., Young, D., Byrne, K., & Hoyle, R. (2022). Cognitive empathy across the lifespan. Developmental Medicine & Child Neurology.

Eagly, A. H., & Wood, W. (2012). Social role theory. Handbook of theories of social psychology, 2, 458-476.

Fatima, M., & Babu, N. (2023). Cognitive and Affective Empathy in Autism Spectrum Disorders: A Meta-analysis. *Review Journal of Autism and Developmental Disorders*, 1-20.

Garavan, H., Bartsch, H., Conway, K., Decastro, A., Goldstein, R. Z., Heeringa, S., ... & Zahs, D. (2018). Recruiting the ABCD sample: Design considerations and procedures. *Developmental cognitive neuroscience*, *32*, 16-22.

Gutierrez, B. C., Halim, M. L. D., Martinez, M. A., & Arredondo, M. (2020). The heroes and the helpless: The development of benevolent sexism in children. Sex Roles, 82, 558-569. doi:10.1007/s11199-019-01074-4

Heeringa, S., & Berglund, P. (2020). A guide for population-based analysis of the adolescent brain cognitive development (ABCD) study baseline data. bioRxiv 2020.02. 10.942011.

Heidari, S., Babor, T. F., De Castro, P., Tort, S., & Curno, M. (2016). Sex and gender equity in research: Rationale for the SAGER guidelines and recommended use. Research Integrity and Peer Review, 1, 2. doi:10.1186/s41073-016-0007-6

Hus, V., Bishop, S., Gotham, K., Huerta, M., & Lord, C. (2013). Factors influencing scores on the social responsiveness scale. *Journal of Child Psychology and Psychiatry*, *54*(2), 216-224.

Kaat, A. J., Croen, L. A., Constantino, J., Newshaffer, C. J., & Lyall, K. (2023). Modifying the social responsiveness scale for adaptive administration. *Quality of Life Research*, *32*(8), 2353-2360.

Kaufman, J., Birmaher, B., Brent, D., Rao, U. M. A., Flynn, C., Moreci, P., ... & Ryan, N. (1997). Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. Journal of the American Academy of Child & Adolescent Psychiatry, 36(7), 980-988.

Knickmeyer, R., Baron-Cohen, S., Raggatt, P., & Taylor, K. (2005). Foetal testosterone, social relationships, and restricted interests in children. *Journal of child psychology and psychiatry*, *46*(2), 198-210.

Kung, K. T., Louie, K., Spencer, D., & Hines, M. (2024). Prenatal androgen exposure and sex-typical play behaviour: A meta-analysis of classic congenital adrenal hyperplasia studies. *Neuroscience & Biobehavioral Reviews*, 105616.

Leung, R. C., Vogan, V. M., Powell, T. L., Anagnostou, E., & Taylor, M. J. (2016). The role of executive functions in social impairment in Autism Spectrum Disorder. *Child Neuropsychology*, *22*(3), 336-344.

Lyall, K., Constantino, J. N., Weisskopf, M. G., Roberts, A. L., Ascherio, A., & Santangelo, S. L. (2014). Parental social responsiveness and risk of autism spectrum disorder in offspring. *JAMA psychiatry*, *71*(8), 936-942.

Matthews, K. A., & Gallo, L. C. (2011). Psychological perspectives on pathways linking socioeconomic status and physical health. *Annual review of psychology*, *62*, 501-530.

McDonald, S. (2014). Special series on the Cognition Battery of the NIH Toolbox. *J Int Neuropsychol Soc*, *20*, 487-651.

Miller, D. I., & Halpern, D. F. (2014). The new science of cognitive sex differences. *Trends in cognitive sciences*, *18*(1), 37-45.

O'Rourke, S., Whalley, H., Janes, S., MacSweeney, N., Skrenes, A., Crowson, S., ... & Schwannauer, M. (2020). The development of cognitive and emotional maturity in adolescents and its relevance in judicial contexts. *Scottish Sentencing Counc*.

Pallant, J. (2020). SPSS survival manual: A step by step guide to data analysis using IBM SPSS. Routledge.

Reiersen, A. M., Constantino, J. N., Grimmer, M., Martin, N. G., & Todd, R. D. (2008). Evidence for shared genetic influences on self-reported ADHD and autistic symptoms in young adult Australian twins. *Twin Research and Human Genetics*, *11*(6), 579-585.

Research domain Criteria Initiative. (2018). National Institute of Mental Health. https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/units/paradigms/151291

Sedgewick, F., Hill, V., & Pellicano, E. (2019). 'It's different for girls': Gender differences in the friendships and conflict of autistic and neurotypical adolescents. *Autism*, *23*(5), 1119-1132.

Sharp, T. H., Elsabbagh, M., Pickles, A., & Bedford, R. (2023). The subcortical correlates of autistic traits in school-age children: a population-based neuroimaging study. *Molecular Autism*, *14*(1), 1-12.

Silke, C., Brady, B., Boylan, C., & Dolan, P. (2018). Factors influencing the development of empathy and pro-social behaviour among adolescents: A systematic review. *Children and Youth Services Review*, *94*, 421-436.

Stover, P. J., Harlan, W. R., Hammond, J. A., Hendershot, T., & Hamilton, C. M. (2010). PhenX: a toolkit for interdisciplinary genetics research. *Current opinion in lipidology*, *21*(2), 136.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). A Compendium of neuropsychological Tests. NY: Oxford University Press.

Torske, T., Nærland, T., Øie, M. G., Stenberg, N., & Andreassen, O. A. (2018). Metacognitive aspects of executive function are highly associated with social functioning on parent-rated measures in children with autism spectrum disorder. *Frontiers in Behavioral Neuroscience*, *11*, 258.

Waite-Jones, J. M., & Rodriguez, A. M. (2022). 'Medicalisation'and Surveillance of Children and Adolescents' Lives. In *Psychosocial Approaches to Child and Adolescent Health and Wellbeing* (pp. 171-194). Cham: Springer International Publishing.

Wallace, G. L., Dudley, K., Anthony, L., Pugliese, C. E., Orionzi, B., Clasen, L., ... & Kenworthy, L. (2017). Divergence of age-related differences in social-communication: Improvements for typically developing youth but declines for youth with autism spectrum disorder. *Journal of autism and developmental disorders*, *47*, 472-479.

Wechsler, D. (2014). *Wechsler intelligence scale for children: Grundl Testauswertung und Interpretation/[Übers. durch Pearson Assessment & Information GmbH, Frankfurt/M.]. Manual 1*. Pearson.

Weintraub, S., Bauer, P. J., Zelazo, P. D., Wallner-Allen, K., Dikmen, S. S., Heaton, R. K., ... & Gershon, R. C. (2013). I. NIH Toolbox Cognition Battery (CB): introduction and pediatric data. *Monographs of the Society for Research in Child Development*, *78*(4), 1-15.

Whitlock, A., Fulton, K., Lai, M. C., Pellicano, E., & Mandy, W. (2020). Recognition of girls on the autism spectrum by primary school educators: An experimental study. *Autism Research*, *13*(8), 1358-1372.

Wood-Downie, H., Wong, B., Kovshoff, H., Cortese, S., & Hadwin, J. A. (2021). Research Review: A systematic review and meta-analysis of sex/gender differences in social interaction and communication in autistic and nonautistic children and adolescents. *Journal of Child Psychology and Psychiatry*, *62*(8), 922-936

Appendices

Appendix A. PRISMA Reporting Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Pg 8
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Pg 9
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Pg 12,13,14
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Pg 15
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Pg 19,20
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Pg 17
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Pg 16,17,18

Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Pg18, 19, 20,22
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Pg 20,21
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Pg 16,17
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Pg 17
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Pg 21,22
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	n/a
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Pg 21
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	n/a
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Pg 21
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Pg 21, 28-33
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	n/a
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	n/a
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	n/a

Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Pg19
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Pg 24
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Pg 19
Study characteristics	17	Cite each included study and present its characteristics.	Pg 28-29
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Pg 25
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Pg30-33
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Pg 32-34
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Pg37
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Pg 30
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	n/a
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	n/a
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Pg 25
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Pg 38-40
	23b	Discuss any limitations of the evidence included in the review.	Pg 41

	23c	Discuss any limitations of the review processes used.	Pg 41
	23d	Discuss implications of the results for practice, policy, and future research.	Pg 42,43
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Pg 16
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Pg 16
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	n/a
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Pg 44
Competing interests	26	Declare any competing interests of review authors.	Pg 44
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	n/a

Appendix B. Search Strategies

Appendix B1. Search Strategy for PubMed

1. ((("cyberbullying"[MeSH Terms]
2. OR ("cyberbull*"[Title/Abstract] OR "cyber bull*"[Title/Abstract])
3. OR ("cyber"[Title/Abstract] OR "internet"[Title/Abstract] OR "digital"[Title/Abstract] OR "online"[Title/Abstract] OR "virtual"[Title/Abstract] OR ((("electronical"[All Fields] OR "electronically"[All Fields] OR "electronics"[MeSH Terms] OR "electronics"[All Fields] OR "electronic"[All Fields]) AND "N2"[All Fields]) AND "bull*"[Title/Abstract]) OR "bull*"[Title/Abstract] OR "harass*"[Title/Abstract] OR ("agress*or"[All Fields] AND "victim*"[Title/Abstract])))
4. AND ("bystand*"[Title/Abstract] OR "cyber bystand*"[Title/Abstract] OR "cyberbystand*"[Title/Abstract] OR "cyber def*"[Title/Abstract] OR "cyberdefend*"[Title/Abstract]))
5. AND ("child*"[Title/Abstract] OR "teen*"[Title/Abstract] OR "adolescen*"[Title/Abstract] OR "young*"[Title/Abstract] OR "school*"[Title/Abstract] OR "minor*"[Title/Abstract] OR "boy"[Title/Abstract] OR "girl*"[Title/Abstract] OR "student*"[Title/Abstract] OR "youth*"[Title/Abstract] OR "juvenile*"[Title/Abstract] OR "kid"[Title/Abstract] OR "underage*"[Title/Abstract])

Appendix B2. Search Strategy for SCOPUS

1. All Fields ‘Cyberbullying’
2. OR
3. AT, AB, KW ‘cyberbull* OR cyber-bull*
4. OR
5. AT, AB, KW (cyber OR internet OR digital OR online OR virtual OR electronic) W/2 (bull* OR harass* OR agress* OR victim)
6. AND
7. All Fields ‘Cyberbystander’
8. OR
9. AT, AB,KW (bystand* OR "cyber-bystand*" OR cyberbystand* OR "cyber-def*" OR cyberdefend*)
10. AND
11. AT, AB, KW (child* OR teen* OR adolescen* OR young* OR school* OR minor* OR boy* OR girl* OR student* OR youth* OR juvenile* OR kid* OR underage* OR preadol*)

Appendix B3. Search Strategy for Web of Science Core Collection (via Web of Knowledge)

1. (TS=(cyberbullying OR "cyber-bullying" OR cyberbull* OR "cyber-bull*") OR
2. TI=(cyberbull* OR "cyber-bull*") OR
3. AB=(cyberbull* OR "cyber-bull*") OR
4. TI=((cyber OR internet OR digital OR online OR virtual OR electronic) ~2 (bull* OR harass* OR agress* OR victim*)) OR
5. AB=((cyber OR internet OR digital OR online OR virtual OR electronic) ~2 (bull* OR harass* OR agress* OR victim*)))
6. AND
7. (TS=(bystander) OR
8. TI=(bystand* OR "cyber-bystand*" OR cyberbystand* OR "cyber-def*" OR cyberdefend*) OR
9. AB=(bystand* OR "cyber-bystand*" OR cyberbystand* OR "cyber-def*" OR cyberdefend*))
10. AND
11. (TI=(child* OR teen* OR adolescen* OR young* OR school* OR minor* OR boy* OR girl* OR student* OR youth* OR juvenile* OR kid* OR underage* OR preadol*) OR
12. AB=(child* OR teen* OR adolescen* OR young* OR school* OR minor* OR boy* OR girl* OR student* OR youth* OR juvenile* OR kid* OR underage* OR preadol*))

Appendix B4. Search Strategy for Psychology and Behavioural Science Collection (via EBESCO host)

1. DE "Cyberbullying"
2. TI (cyberbull* or cyber-bull*) OR AB (cyberbull* or cyber-bull*)
3. TI ((cyber or internet or digital or online or virtual or electronic) N2 (bull* or -bull* or harass* or agress* victim*)) OR AB ((cyber or internet or digital or online or virtual or electronic) N2 (bull* or -bull* or harass* or agress* victim*))
4. S1 OR S2 OR S3
5. Bystander Effect
6. Bystander
7. Bystander Intervention
8. TI (bystand* or cyber-bystand* or cyberbystand* or cyber-def* or cyberdefend*) OR AB (bystand* or cyber-bystand* or cyberbystand* or cyber-def* or cyberdefend*)
9. S5 or S6 or s7 or s8
10. TI (child* or teen* or adolescen* or young* or school* or minor* or boy* or girl* or student* or youth* or juvenile* or kid* or underage* or preadol*) OR (child* or teen* or adolescen* or young* or school* or minor* or boy* or girl* or student* or youth* or juvenile* or kid* or underage* or preadol*)
11. S4 AND S9 AND S10

Appendix C. Descriptive Statistics

Appendix C.1 Sores on Psychological Variables at Baseline.

Variable: Number of completed data (Missing data)	Mean (Standard Deviation)
NIH Picture Vocabulary N=9677 (127)	107.06 (17.19)
NIH Flanker N=9671 (133)	95.43 (13.79)
NIH List Sorting N=9643 (161)	100.64 (14.88)
NIH Card Sorting N=9673 (131)	96.61 (15.17)
NIH Pattern Comparison N=9655 (149)	93.57 (22.25)
NIH Picture Sequence N=9669 (135)	101.00 (16.13)
NIH Reading Recognition N=9665 (139)	103.00 (19.34)
NIH Total Cognitive Score N=9480 (324)	100.60 (18.22)
RAVLT Immediate z-score N=9622 (182)	-.16 (.93)
RAVLT Delayed z-score N=9572 (232)	-.15 (.94)
Matrix Reasoning (WISC) N=9588 (216)	9.86 (3.0)
CBCL Internalizing T-score N=9797 (7)	48.73 (10.67)
CBCL Externalising T-score N=9797 (7)	45.89 (10.32)

Abbreviations: CBCL, The Child Behavior Checklist; NIH, National Institutes of Health; RAVLT, Rey Auditory Verbal Learning Test; WISC, Wechsler Intelligence Scale for Children

Appendix D. ANOVA Summary Tables

Appendix D1. Two-Way Between Groups ANOVA Summary Table Excluding Participants with ASD and/or ADHD.

	SS	*df*	*MS*	*F*	*p*
Age	34.05	3	11.35	.79	.497
Sex	278.81	1	278.81	19.49	<.001
Age*Sex	37.02	3	12.34	.86	.460
Error	129014.38	9017	14.31		

Appendix D2: Two-Way Between Groups ANOVA Summary Table Using Population Weighting.

	df	*F*	*p*
Age	3	.91	.454
Sex	1	21.42	<.001
Age*Sex	3	.91	.709

Appendix D3: Two-Way Between Groups ANOVA Summary Table Using Population Weighting, Excluding Those with ASD and ADHD.

	df	*F*	*p*
Age	3	.63	.606
Sex	1	19.37	<.001
Age*Sex	3	.18	.907

Appendix E. Multiple Regression Models

E1: Multiple Regression Excluding Those with ASD and/or ADHD Summary Table

Multiple Regression Analysis Summary Predicting Social Responsiveness Scores excluding those with ASD and/or ADHD.

Predictor Variable	*B* (un standardised)	Confidence Interval (95%)		t	Sig.	Partial eta-squared
Intercept	17.281	15.707,	18.854	21.531	**<.001**	.056
Sex (male)	.692	.527,	.857	8.222	**<.001**	.009
Born in US (no)	-.398	-.870,	.075	-1.651	.099	.000
NIH Picture Vocabulary	.013	.007,	.018	4.160	**<.001**	.002
NIH Flanker	-.012	-.019,	-.006	-3.556	**<.001**	.002
NIH List Sorting	-.008	-.014,	-.001	-2.301	**.021**	.001
NIH Card Sorting	-.002	-.009,	.004	-.694	.488	.000
NIH Pattern Comparison	-.002	-.006,	.003	-.749	.454	.000
NIH Picture Sequence	-.003	-.008,	.003	-.895	.371	.000
NIH Oral Reading	-.007	-.012,	-.002	-2.772	**.006**	.001
RAVLT IM	-.243	-.393,	-.092	-3.153	**.002**	.001
RAVLT DL	.028	-.125,	.180	.355	.722	.000
WISC Matrix Reasoning	.010	-.021,	.042	.636	.525	.000
Household Income	-.201	-.239,	-.163	-10.420	**<.001**	.014
Age	-.009	-.141,	.124	-.128	.898	.000

Adjusted R^2 =4.1 percent; (F (14, 7843) =25.01; p<.001). Sample is unrelated participants excluding those with ASD/ADHD diagnoses.

Abbreviations: NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

E2: Multiple Regression Summary Table Using Population Weighting

Multiple Regression Analysis Summary Predicting Social Responsiveness Using Population Weighted Samples

Predictor Variable	Estimate	Confidence Interval (95%)		t	Sig.
(Intercept)	17.339	15.750,	18.928	22.760	**<.001**
Sex (male)	1.002	.741,	1.264	7.997	**<.001**
Born in the US (no)	-.736	-1.199,	-.273	-3.318	**.003**
NIH Picture Vocabulary	.014	.004,	.025	2.939	**.008**
NIH Flanker	-.020	-.028,	-.013	-5.562	**<.001**
NIH List Sorting	-.007	-.017,	.002	-1.627	.119
NIH Card Sorting	-.004	-.012,	.004	-.983	.337
NIH Pattern Comparison	.001	-.007,	.008	.176	.862
NIH Picture Sequence	.000	-.007,	.007	.048	.962
NIH Oral Reading	-.011	-.020,	-.002	-2.531	**.020**
RAVLT IM	-.273	-.527,	-.018	-2.236	**.037**
RAVLT DL	-.008	-.221,	.205	-.079	.938
WISC Matrix Reasoning	.034	-.016,	.084	1.407	.175
Household Income	-.226	-.301,	-.151	-6.266	**<.001**
Age	.037	-.151,	.225	.413	.684

Adjusted R^2 =5.5 percent. Sample is population weighted unrelated including those with ASD/ADHD diagnoses.

Abbreviations: NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

E3: Multiple Regression Summary Table Using Population Weighting Excluding Participants with ADHD and ASD

Multiple Regression Analysis Summary Predicting Social Responsiveness Using Population Weighted Samples Excluding Participants with ADHD and ASD

Predictor Variable	Estimate	Confidence Interval (95%)		t	Sig.
(Intercept)	17.915	16.675,	19.156	30.125	**<.001**
Sex (male)	.725	.488,	.962	6.383	**<.001**
Born in the US (no)	-.577	-1.028,	-.126	-2.670	**.015**
NIH Picture Vocabulary	.012	.002,	.022	2.570	**.018**
NIH Flanker	-.015	-.021,	-.009	-5.092	**<.001**
NIH List Sorting	-.008	-.016,	-.001	-2.311	**.032**
NIH Card Sorting	.004	-.012,	.004	-1.005	.327
NIH Pattern Comparison	.002	-.004,	.008	.677	.506
NIH Picture Sequence	-.002	-.009,	.005	-.654	.521
NIH Oral Reading	-.011	-.021,	-.002	-2.539	**.020**
RAVLT IM	-.222	-.469,	.025	-1.874	.076
RAVLT DL	.005	-.238,	.227	-.049	.962
WISC Matrix Reas	.018	-.027,	.064	.843	.409
Household Income	-.221	-.297,	-.146	-6.12	**<.001**
Age	-.018	-.180,	.144	-.235	.817

Adjusted R^2 =5.1 percent. Sample is population weighted unrelated participants excluding those with ASD/ADHD diagnoses.

Abbreviations: NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

Appendix F. Male Multiple Regression Models

F1: Male Multiple Regression Summary Table Excluding Participants with ADHD and ASD

Male Multiple Regression Analysis Summary Predicting Social Responsiveness Excluding Participants with ADHD and ASD

Predictor Variable	*B* (un standardised	Confidence Interval (95%)		t	Sig	Partial Eta-Squared
Intercept	19.803	17.444,	22.163	16.454	**<.001**	.063
Born in US (no)	-.811	-1.540,	-.082	-2.180	**.029**	.001
NIH Picture Vocabulary	.015	.006,	.024	3.315	**<.001**	.003
NIH Flanker	-.015	-.025,	-.005	-3.028	**.002**	.002
NIH List Sorting	-.012	-.022,	-.002	-2.349	**.019**	.001
NIH Card Sorting	-.004	-.013,	.006	-.753	.452	.000
NIH Pattern Comparison	.002	-.004,	.008	.728	.467	.000
NIH Picture Sequence	-.004	-.013,	.005	-.931	.352	.000
NIH Oral Reading	-.007	-.015,	.000	-1.910	.056	.001
RAVLT IM	-.278	-.504,	-.052	-2.408	**.016**	.001
RAVLT DL	.006	-.222,	.233	.049	.961	.000
WISC Matrix Reasoning	.047	.000,	.094	1.967	**.049**	.001
Household Income	-.230	-.286,	-.174	-8.056	**<.001**	.016
Age	-.173	-.371,	.024	-1.721	.085	.001

Adjusted R^2 =3.6 percent; (F (13, 4060) =12.651; $p<.001$). Sample is unrelated males excluding participants with ASD and/or ADHD.

Abbreviations: NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

F2: Male Multiple Regression Using Population Weighting

Male Multiple Regression Analysis Summary Predicting Social Responsiveness Using Population Weighted Samples

Predictor Variable	Estimate	Confidence Interval (95%)		t	Sig.
(Intercept)	18.833	15.435,	22.230	11.562	**<.001**
Born in the US (no)	-1.317	-1.906,	-.728	-4.661	**<.001**
NIH Picture Vocabulary	.011	-.004,	.025	1.558	.135
NIH Flanker	-.026	-.042,	-.010	-3.345	**.003**
NIH List Sorting	-.010	-.025,	.005	-1.397	.178
NIH Card Sorting	-.004	-.017,	.009	-.619	.543
NIH Pattern Comparison	.004	-.007,	.015	.731	.473
NIH Picture Sequence	.001	-.010,	.013	.275	.786
NIH Oral Reading	-.010	-.023,	.003	-1.667	.111
RAVLT IM	-.363	-.698,	-.029	-2.266	**.035**
RAVLT DL	-.045	-.365,	.274	-.035	.770
WISC Matrix Reasoning	.108	.030,	.186	2.266	**.009**
Household Income	-.262	-.363,	-.161	-5.400	**<.001**
Age	.008	-.289,	.306	.058	.953

Adjusted R^2 =4.6 percent. Sample is populated weighted male including those with ASD/ADHD diagnoses.

Abbreviations: NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

F3: Male Multiple Regression Summary Table Using Population Weighting Excluding Participants with ADHD and ASD

Male Multiple Regression Analysis Summary Predicting Social Responsiveness Using Population Weighting Excluding Participants with ADHD and ASD

Predictor Variable	Estimate	Confidence Interval (95%)		t	Sig.
(Intercept)	19.609	17.012,	22.206	15.749	**<.001**
Born in the US (no)	-1.032	-1.614,	-.449	-3.691	**.001**
NIH Picture Vocabulary	.007	-.005,	.019	1.236	.231
NIH Flanker	-.016	-.031,	-.002	-2.369	**.028**
NIH List Sorting	-.011	-.022,	-8.157	-2.101	**.049**
NIH Card Sorting	-.004	-.016,	.008	-0.638	.531
NIH Pattern Comparison	.007	-.003,	.017	1.379	.183
NIH Picture Sequence	-.002	-.013,	.008	-.448	.659
NIH Oral Reading	-.011	-.024,	.001	-1.860	.078
RAVLT IM	-.275	-.586,	.036	-1.845	.080
RAVLT DL	-.046	-.376,	.284	-.293	.773
WISC Matrix Reasoning	.078	9.637,	.156	2.086	**.050**
Household Income	-.248	-.354,	-.142	-4.878	**<.001**
Age	-.10	-.371,	.178	-0.73	.472

Adjusted R^2 =4.8 percent. Sample is population weighted unrelated males excluding those with ASD/ADHD diagnoses.

Abbreviations: NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

Appendix G. Female Regression Models

G1: Female Multiple Regression Summary Table Excluding Those with ASD and/or ADHD

Female Multiple Regression Analysis Summary Predicting Social Responsiveness Excluding Participants with ADHD and ASD

Predictor Variable	*B* (un standardised	95% Confidence Interval		t	Sig.	Partial Eta-Squared
Intercept	15.295	13.236,	17.354	14.562	**<.001**	.053
Born in US (no)	.008	-.591,	.607	.027	.979	.000
NIH Picture Vocabulary	.010	.002,	.018	2.595	**.010**	.002
NIH Flanker	-.009	-.018,	.001	-1.811	.070	.001
NIH List Sorting	-.004	-.013,	.004	-.947	.344	.000
NIH Card Sorting	.000	-.009,	.008	-.073	.942	.000
NIH Pattern Comparison	-.007	-.012,	-.001	-2.272	**.023**	.001
NIH Picture Sequence	-.001	-.009,	.006	-.306	.760	.000
NIH Oral Reading	-.007	-.013,	.000	-2.021	**.043**	.001
RAVLT IM	-.206	-.403,	-.009	-2.051	**.040**	.001
RAVLT DL	.058	-.142,	.258	.567	.571	.000
WISC Matrix Reas	-.028	-.069,	.013	-1.329	.184	.000
Household Income	-.170	-.220,	-.120	-6.664	**<.001**	.012
Age	.172	-.002,	.346	1.940	.052	.001

Adjusted R^2 =3 percent; (F (13, 3770) =10.012; $p<.001$). Sample is unrelated females excluding participants with ASD and/or ADHD.

Abbreviations: NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

G2: Female Multiple Regression Using Population Weighting

Female Multiple Regression Analysis Summary Predicting Social Responsiveness Using Population Weighted Samples

Predictor Variable	Estimate	Confidence Interval (95%)		t	Sig.
(Intercept)	17.046	14.763,	19.329	15.572	**<.001**
Born in the US (no)	-.170	-.934,	.594	-.463	.648
NIH Picture Vocabulary	.018	.002,	.033	2.418	**.025**
NIH Flanker	-.014	-.028,	.001	-1.966	.063
NIH List Sorting	-.006	-.016,	.005	-1.109	.280
NIH Card Sorting	-.003	-.015,	.008	-.573	.573
NIH Pattern Comparison	-.003	-.011,	.004	-.924	.367
NIH Picture Sequence	-.002	-.010,	.006	-.465	.647
NIH Oral Reading	-.011	-.020,	-.001	-2.359	**.029**
RAVLT IM	-.181	-.507,	.144	-1.164	.258
RAVLT DL	.037	-.222,	.297	.301	.767
WISC Matrix Reasoning	-.048	-.129,	.032	-1.252	.225
Household Income	-.184	-.270,	-.098	-4.450	**<.001**
Age	.050	-.209,	.310	.406	.689

Adjusted R^2 =4.2 percent. Sample is population weighted females including those with ASD/ADHD diagnoses.

Abbreviations: NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

G3: Female Multiple Regression Using Population Weighting Excluding Participants with ADHD and ASD

Female Multiple Regression Analysis Summary Predicting Social Responsiveness Using Population Weighted Samples Excluding Participants with ASD and ADHD

Predictor Variable	Estimate	Confidence Intervals (95%)		t	Sig.
(Intercept)	17.099	15.006,	19.192	17.039	**<.001**
Born in the US (no)	-.145	-.914,	.625	-.392	.699
NIH Picture Vocabulary	.017	.002,	.033	2.325	**.031**
NIH Flanker	-.013	-.028,	.002	-1.760	.094
NIH List Sorting	-.006	-.016,	.004	-1.274	.217
NIH Card Sorting	-.003	-.015,	.008	-.580	.569
NIH Pattern Comparison	-.003	-.011,	.004	-.951	.353
NIH Picture Sequence	-.002	-.010,	.006	-.608	.550
NIH Oral Reading	-.010	-.020,	-.001	-2.294	**.033**
RAVLT IM	-.168	-.505,	.168	-1.043	.309
RAVLT DL	.041	-.223,	.305	.326	.748
WISC Matrix Reasoning	-.045	-.127,	.037	-1.134	.270
Household Income	-.192	-.275,	-.108	-4.794	**<.001**
Age	.053	-.200,	.307	.440	.665

Adjusted R^2 =4.3 percent. Sample is population weighted unrelated females excluding those with ASD/ADHD diagnoses.

Abbreviations: NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

Appendix H. Multiple Regression Models with CBCL Scores

Appendix H1. Multiple Regression Analysis with CBCL Scores.

Multiple Regression Analysis Summary Table for Social Responsiveness Scores including CBCL Scores

Predictor Variable	*B*(unstandardised)	Confidence Interval (95%)		t	Sig.	Partial Eta Squared
Intercept	5.878	4.293,	7.464	7.269	**<.001**	.007
Sex (male)	.655	.496,	.813	8.093	**<.001**	.008
Born in US (no)	-.408	-.865,	.048	-1.754	.080	.000
NIH Picture Vocabulary	.009	.003,	.015	3.176	**.002**	.001
NIH Flanker	-.013	-.020,	-.007	-3.984	**<.001**	.002
NIH List Sorting	-.007	-.013,	.000	-2.096	**.036**	.001
NIH Card Sorting	-.001	-.007,	.005	-.360	.719	.000
NIH Pattern Comparison	-.002	-.006,	.002	-1.012	.311	.000
NIH Picture Sequence	.003	-.003,	.008	.919	.358	.000
NIH Oral Reading	-.006	-.011,	-.001	-2.336	**.020**	.001
RAVLT IM	-.270	-.414,	-.126	-3.668	**<.001**	.002
RAVLT DL	.037	-.108,	.183	.504	.615	.000
WISC Matrix Reasoning	.026	-.004,	.056	1.671	.095	.000
CBCL Internalising	.118	.108,	.127	25.161	**<.001**	.073
CBCL Externalising	.075	.066,	.085	15.431	**<.001**	.029
Household income	-.106	-.142,	-.070	-5.706	**<.001**	.004
Age	.112	-.015,	.239	1.731	.083	.000

Adjusted R^2 =23.6 percent; (F (16, 8012=155.60, p<.001). Sample is unrelated participants including those with ASD/ADHD diagnoses.

Abbreviations: CBCL, Child Behaviour Checklist; NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

H2: Multiple Regression with CBCL Scores Excluding Those with ASD and/or ADHD Summary Table

Multiple Regression with CBCL Scores Analysis Summary Predicting Social Responsiveness Scores excluding those with ASD and/or ADHD.

Predictor Variable	*B*(unstandardised)	Confidence Interval (95%)		t	Sig.	Partial Eta Squared
Intercept	6.996	5.517,	8.474	9.274	**<.001**	.011
Sex (male)	.447	.300,	.595	5.950	**<.001**	.004
Born in US (no)	-.288	-.709,	.133	-1.339	.181	.000
NIH Picture Vocabulary	.007	.002,	.013	2.766	**.006**	.001
NIH Flanker	-.010	-.016,	-.003	-3.037	**.002**	.001
NIH List Sorting	-.006	-.012,	.000	-1.909	.056	.000
NIH Card Sorting	-.001	-.006,	.005	-.177	.860	.000
NIH Pattern Comparison	-.001	-.005,	.002	-.706	.480	.000
NIH Picture Sequence	.001	-.005,	.006	.201	.841	.000
NIH Oral Reading	-.007	-.012,	-.003	-3.252	**.001**	.001
RAVLT IM	-.235	-.370,	-.101	-3.430	**<.001**	.001
RAVLT DL	.058	-.078,	.193	.829	.407	.000
WISC Matrix Reas	.018	-.010,	.046	1.248	.212	.000
CBCL Internalizing	.110	.101,	.118	25.195	**<.001**	.075
CBCL Externalising	.072	.063,	.081	15.857	**<.001**	.031
Household income	-.110	-.144,	-.077	-6.383	**<.001**	.005
Age	.056	-.062,	.174	.927	.354	.000

Adjusted R^2 =23.9 percent; (F (16, 7840=154.80, p<.001). Sample is unrelated participants excluding those with ASD/ADHD diagnoses.

Abbreviations: CBCL, Child Behaviour Checklist; NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

H3: Multiple Regression with CBCL Scores Using Population Weighting

Multiple Regression with CBCL Scores Analysis Summary Predicting Social Responsiveness Scores Using Population Weighted Samples

Predictor Variable	Estimate	Confidence Interval (95%)		t	Sig.
(Intercept)	6.628	4.875,	8.381	7.887	**<.001**
Sex (male)	.730	.465,	.994	5.754	**<.001**
Born in the US (no)	-.547	-1.065	-.028	-2.199	**.040**
NIH Picture Vocabulary	.006	-.003,	.015	1.402	.176
NIH Flanker	-.016	-.023,	-.009	-4.876	**<.001**
NIH List Sorting	-.005	-.015,	.005	-.980	.339
NIH Card Sorting	-.002	-.009,	.005	-.709	.486
NIH Pattern Comparison	.001	-.006,	.007	.170	.866
NIH Picture Sequence	.002	-.004,	.009	.798	.434
NIH Oral Reading	-.009	-.017,	-.002	-2.508	**.021**
RAVLT IM	-.276	-.486,	-.066	-2.747	**.012**
RAVLT DL	.067	-.125,	.259	.728	.475
WISC Matrix Reasoning	.040	-.002,	.083	1.980	.062
CBCL Internalising	.125	.111,	.138	19.335	**<.001**
CBCL Externalising	.071	.053,	.089	8.296	**<.001**
Household income	-.117	-.175,	-.059	-4.207	**<.001**
Age	.070	-.073,	.213	1.024	.318

Adjusted R^2 =25.8 percent. Sample is population weighted unrelated participants including those with ASD/ADHD diagnoses

Abbreviations: CBCL, Child Behaviour Checklist; NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

H4: Multiple Regression with CBCL Scores Using Population Weighting Excluding Participants with ADHD and ASD

Multiple Regression with CBCL Scores Analysis Summary Predicting Social Responsiveness Scores Using Population Weighted Samples Excluding Participants with ADHD and ASD.

Predictor Variable	Estimate	Confidence Intervals (95%)		t	Sig.
(Intercept)	7.664	6.276,	9.052	11.519	**<.001**
Sex (male)	.494	.250,	.738	4.219	**<.001**
Born in the US (no)	-.410	-.919,	.099	-1.682	.108
NIH Picture Vocabulary	.005	-.004,	.014	1.175	.254
NIH Flanker	-.011	-.018,	-.005	-3.883	**<.001**
NIH List Sorting	-.005	-.013,	.003	-1.328	.199
NIH Card Sorting	-.003	-.009,	.004	-.824	.419
NIH Pattern Comparison	.002	-.004,	.007	.728	.475
NIH Picture Sequence	-5.952	-.005,	.005	-.023	.982
NIH Oral Reading	-.010	-.018,	-.002	-2.707	**.014**
RAVLT IM	-.232	-.440,	-.024	-2.331	**.030**
RAVLT DL	.070	-.134,	.274	.715	.483
WISC Matrix Reasoning	.025	-.016,	.067	1.268	.219
CBCL Internalizing	.118	.104,	.131	18.208	**<.001**
CBCL Externalising	.068	.050,	.086	7.841	**<.001**
Household income	-.118	-.178,	-.058	-4.099	**<.001**
Age	.022	-.091,	.135	.409	.687

Adjusted R^2 =26.1 percent. Sample is population weighted unrelated participants excluding those with ASD/ADHD diagnoses

Abbreviations: CBCL, Child Behaviour Checklist; NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

Appendix I. Male Multiple Regression Models with CBCL Scores

Appendix I1. Male Regression Model with CBCL Scores

Multiple Regression Analysis Summary Table Predicting Social Responsiveness Scores in Males with CBCL Scores

Predictor Variable	*B*(unstandardised)	Confidence Interval (95%)		t	Sig.	Partial Eta Squared
Intercept	6.333	3.921,	8.746	5.147	**<.001**	.006
Born in US (no)	-1.035	-1.753,	-.317	-2.825	**.005**	.002
NIH Picture Vocabulary	.014	0.006,	.023	3.203	**.001**	.002
NIH Flanker	-.017	-0.027,	-.008	-3.498	**<.001**	.003
NIH List Sorting	-.010	-0.020,	-.001	-2.093	**.036**	.001
NIH Card Sorting	-.004	-0.013,	.005	-.805	.421	.000
NIH Pattern Comparison	.000	-0.006,	.006	-.078	.938	.000
NIH Picture Sequence	.002	-0.007,	.010	.366	.715	.000
NIH Oral Reading	-.004	-0.011,	.003	-1.121	.262	.000
RAVLT IM	-.287	-0.507,	-.067	-2.561	**.010**	.002
RAVLT DL	-.010	-0.231,	.211	-.089	.929	.000
WISC Matrix Reasoning	.056	0.010,	.101	2.405	**.016**	.001
CBCL Internalising	.140	0.126,	.153	19.919	**<.001**	.086
CBCL Externalising	.075	0.061,	.089	10.503	**<.001**	.026
Household income	-.120	-0.175,	-.066	-4.312	**<.001**	.004
Age	.025	-0.167,	.217	.258	.797	.000

Adjusted R^2 =23.8 percent; (*F* (15, 4198=88.83, *p*<.001). *Sample is unrelated male participants including those with ASD/ADHD diagnoses.*

Abbreviations: CBCL, Child Behaviour Checklist; NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

I2: Male Multiple Regression with CBCL Scores Excluding Participants with ASD/ADHD

Multiple Regression Analysis Summary Predicting Social Responsiveness Scores in Males with CBCL Scores Excluding Participants with ASD/ADHD

Predictor Variable	*B*(Unstandardised)	Confidence Intervals (95%)		t	Sig.	Partial Eta Squared
Intercept	7.987	5.785,	10.189	7.110	<**.001**	.012
Born in US (no)	-.816	-1.460,	-.171	-2.481	**.013**	.002
NIH Picture Vocabulary	.012	.004,	.020	3.041	**.002**	.002
NIH Flanker	-.012	-.021,	-.003	-2.698	**.007**	.002
NIH List Sorting	-.007	-.016,	.002	-1.586	.113	.001
NIH Card Sorting	-.003	-.011,	.006	-.638	.523	.000
NIH Pattern Comparison	.001	-.004,	.007	.544	.586	.000
NIH Picture Sequence	-.002	-.009,	.006	-.405	.686	.000
NIH Oral Reading	-.008	-.015,	-.001	-2.394	**.017**	.001
RAVLT IM	-.244	-.444,	-.045	-2.401	**.016**	.001
RAVLT DL	.009	-.192,	.210	.089	.929	.000
WISC Matrix Reasoning	.041	-.001,	.082	1.934	.053	.001
CBCL Internalizing	.128	.115,	.140	20.009	<**.001**	.090
CBCL Externalising	.070	.058,	.083	10.863	<**.001**	.028
Household income	-.123	-.172,	-.073	-4.819	<**.001**	.006
Age	-.085	-.260,	.089	-.957	.339	.000

Adjusted R^2 *=24.8 percent;* (*F* (15, 4057=90.74, *p*<.001). *Sample is unrelated male participants excluding those with ASD/ADHD diagnoses.*

Abbreviations: CBCL, Child Behaviour Checklist; NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

I3: Multiple Regression with CBCL Scores in Males Using Population Weighting

Multiple Regression with CBCL Scores Analysis Summary Predicting Social Responsiveness Scores in Males Using Population Weighted Samples

Predictor Variable	Estimate	Lower	Upper	t	Sig.
Intercept	6.297	2.728	9.866	3.680	**<.001**
Born in the US (no)	-1.131	-1.708	-.554	-4.089	**<.001**
NIH Picture Vocabulary	.003	-.011	.018	.488	.631
NIH Flanker	-.022	-.036	-.008	-3.258	**.004**
NIH List Sorting	-.005	-.021	.011	-.623	.540
NIH Card Sorting	-.004	-.016	.007	-.775	.447
NIH Pattern Comparison	.002	-.007	.011	.387	.703
NIH Picture Sequence	.004	-.006	.015	.847	.407
NIH Oral Reading	-.008	-.019	.003	-1.576	.131
RAVLT IM	-.298	-.570	-.026	-2.283	**.034**
RAVLT DL	-.011	-.243	.221	-.097	.924
WISC Matrix Reasoning	.091	.029	.153	3.041	**.006**
CBCL Internalising	.152	.132	.171	16.513	**<.001**
CBCL Externalising	.073	.051	.095	6.858	**<.001**
Household income	-.121	-.186	-.056	-3.900	**<.001**
Age	.052	-.213	.317	.410	.686

Adjusted R^2 =27.3 percent. Sample is population weighted male participants including those with ASD/ADHD diagnoses

Abbreviations: CBCL, Child Behaviour Checklist; NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

I4: Multiple Regression with CBCL Scores in Males Using Population Weighting Excluding Those with ADHD and ASD

Multiple Regression with CBCL Scores Analysis Summary Predicting Social Responsiveness Scores Using Population Weighting Excluding Participants with ADHD and ASD.

Predictor Variable	Estimate	Confidence Intervals (95%)		t	Sig.
Intercept	7.831	5.082,	10.581	5.941	**.000**
Born in the US (no)	-.883	-1.464,	-.301	-3.168	**.005**
NIH Picture Vocabulary	.002	-.010,	.013	.302	.766
NIH Flanker	-.014	-.027,	-.001	-2.313	**.032**
NIH List Sorting	-.005	-.017,	.006	-.916	.370
NIH Card Sorting	-.004	-.015,	.006	-.851	.405
NIH Pattern Comparison	.005	-.003,	.013	1.213	.239
NIH Picture Sequence	1.239	-.009,	.009	.003	.998
NIH Oral Reading	-.010	-.021,	.001	-1.966	.063
RAVLT IM	-.228	-.475,	.019	-1.923	.069
RAVLT DL	-.009	-.254,	.235	-.078	.938
WISC Matrix Reasoning	.062	-.003,	.127	2.003	.059
CBCL Internalizing	.139	.124,	.155	18.917	**<.000**
CBCL Externalising	.068	.047,	.089	6.819	**<.000**
Household income	-.118	-.194,	-.042	-3.225	**.004**
Age	-.038	-.274,	.199	-.332	.743

Adjusted R^2 =28 percent. Sample is population weighted unrelated participants excluding those with ASD/ADHD diagnoses

Abbreviations: CBCL, Child Behaviour Checklist; NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

Appendix J. Female Multiple Regression Models with CBCL Scores

J1: Female Multiple Regression with CBCL Scores

Multiple Regression Analysis Summary Table Predicting Social Responsiveness Scores in Females with CBCL Scores

Predictor Variable	*B*(unstandardised)	Confidence Interval (95%)		t	Sig.	Partial Eta Squared
Intercept	6.067	4.064,	8.070	5.938	**<.001**	.009
Born in US (no)	.147	-.408,	.702	.519	.604	.000
NIH Picture Vocabulary	.005	-.002,	.012	1.395	.163	.001
NIH Flanker	-.008	-.017,	.000	-1.862	.063	.001
NIH List Sorting	-.003	-.011,	.005	-.739	.460	.000
NIH Card Sorting	.002	-.006,	.010	.491	.623	.000
NIH Pattern Comparison	-.005	-.010,	.000	-1.858	.063	.001
NIH Picture Sequence	.004	-.003,	.010	1.011	.312	.000
NIH Oral Reading	-.008	-.014,	-.002	-2.474	**.013**	.002
RAVLT IM	-.247	-.428,	-.066	-2.669	**.008**	.002
RAVLT DL	.092	-.093,	.276	.976	.329	.000
WISC Matrix Reasoning	-.010	-.048,	.029	-.493	.622	.000
CBCL Internalising	.094	.082,	.106	15.681	**<.001**	.061
CBCL Externalising	.073	.060,	.086	11.317	**<.001**	.033
Household income	-.091	-.137,	-.044	-3.830	**<.001**	.004
Age	.206	.046,	.367	2.524	**.012**	.002

Adjusted R^2 *=21.3 percent;* (*F* (15, 3799=69.89, *p*<.001). *Sample is unrelated female participants including those with ASD/ADHD diagnoses.*

Abbreviations: CBCL, Child Behaviour Checklist; NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

J2: Multiple Regression with CBCL Scores in Females Excluding Participants with ADHD and ASD

Multiple Regression with CBCL Scores in Females Analysis Summary Predicting Social Responsiveness Scores Excluding Participants with ADHD and ASD.

Predictor Variable	*B*(unstandardised)	Confidence Interval (95%)		t	Sig.	Partial Eta Squared
Intercept	6.357	4.402,	8.312	6.375	**<.001**	.011
Born in US (no)	.184	-.355,	.724	.669	.504	.000
NIH Picture Vocabulary	.003	-.003,	.010	.988	.323	.000
NIH Flanker	-.006	-.015,	.002	-1.423	.155	.001
NIH List Sorting	-.005	-.012,	.003	-1.166	.244	.000
NIH Card Sorting	.002	-.005,	.010	.540	.589	.000
NIH Pattern Comparison	-.005	-.010,	.000	-2.037	**.042**	.001
NIH Picture Sequence	.003	-.004,	.009	.760	.447	.000
NIH Oral Reading	-.007	-.013,	-.001	-2.273	**.023**	.001
RAVLT IM	-.223	-.400,	-.045	-2.463	**.014**	.002
RAVLT DL	.117	-.064,	.297	1.269	.204	.000
WISC Matrix Reas	-.008	-.045,	.030	-.396	.692	.000
CBCL Internalizing	.091	.080,	.102	15.569	**<.001**	.060
CBCL Externalising	.072	.060,	.084	11.425	**<.001**	.033
Household income	-.098	-.143,	-.053	-4.236	**<.001**	.005
Age	.211	.055,	.368	2.644	**.008**	.002

Adjusted R^2 *=21.8 percent;* (*F* (15, 3768=69.85, *p*<.001). *Sample is unrelated female participants excluding those with ASD/ADHD diagnoses.*

Abbreviations: CBCL, Child Behaviour Checklist; NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

J3: Multiple Regression with CBCL Scores in Females Using Population Weighted Samples

Multiple Regression with CBCL Scores Analysis Summary Predicting Social Responsiveness Scores in Females Using Population Weighted Samples

Predictor Variable	Estimate	Confidence Intervals (95%		t	Sig.
Intercept	8.108	5.294,	10.922	6.010	**<.001**
Born in the US (no)	-.010	-.707,	.686	-.031	.975
NIH Picture Vocabulary	.010	-.004,	.023	1.510	.147
NIH Flanker	-.010	-.022,	.003	-1.550	.137
NIH List Sorting	-.005	-.014,	.005	-1.053	.305
NIH Card Sorting	-.001	-.011,	.009	-.170	.867
NIH Pattern Comparison	-.002	-.009,	.005	-.489	.630
NIH Picture Sequence	4.255	-.007,	.007	.012	.991
NIH Oral Reading	-.010	-.018,	-.001	-2.404	**.026**
RAVLT IM	-.239	-.523,	.046	-1.748	.096
RAVLT DL	.145	-.094,	.383	1.264	.221
WISC Matrix Reasoning	-.022	-.087,	.043	-.721	.479
CBCL Internalising	.096	.074,	.117	9.278	**<.001**
CBCL Externalising	.066	.043,	.088	6.063	**<.001**
Household income	-.109	-.188,	-.030	-2.880	**.009**
Age	.078	-.105,	.261	.889	.384

Adjusted R^2 =22.1 percent. Sample is population weighted female participants including those with ASD/ADHD diagnoses

Abbreviations: CBCL, Child Behaviour Checklist; NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

J4: Multiple Regression with CBCL Scores in Females Using Population Weighting Excluding Participants with ADHD and ASD

Multiple Regression with CBCL Scores Analysis Summary Predicting Social Responsiveness Scores in Females Using Population Weighting Excluding Participants with ADHD and ASD.

Predictor Variable	Estimate	Confidence Intervals (95%)		t	Sig.
Intercept	8.230	5.509,	10.951	6.310	**<.001**
Born in the US (no)	.012	-.692,	.716	.036	.972
NIH Picture Vocabulary	.009	-.004,	.022	1.395	.178
NIH Flanker	-.009	-.022,	.004	-1.407	.175
NIH List Sorting	-.005	-.014,	.004	-1.166	.257
NIH Card Sorting	-.001	-.011,	.009	-.185	.855
NIH Pattern Comparison	-.002	-.009,	.005	-.547	.590
NIH Picture Sequence	.000	-.008,	.007	-.113	.911
NIH Oral Reading	-.009	-.018,	-.001	-2.355	**.029**
RAVLT IM	-.226	-.518,	.067	-1.610	.123
RAVLT DL	.150	-.088,	.387	1.317	.203
WISC Matrix Reasoning	-.018	-.085,	.049	-.560	.581
CBCL Internalizing	.095	.072,	.117	8.934	**<.001**
CBCL Externalising	.066	.043,	.088	6.008	**<.001**
Household income	-.117	-.194,	-.040	-3.176	**.005**
Age	.081	-.095,	.257	.958	.349

Adjusted R^2 =22.5 percent. Sample is population weighted female participants excluding those with ASD/ADHD diagnoses

Abbreviations: CBCL, Child Behaviour Checklist; NIH, National Institutes of Health; RAVLT DL, Rey Auditory Verbal Learning Test Delayed Recall; RAVLT IM, Rey Auditory Verbal Learning Test Immediate Recall; US, United States; WISC, Wechsler Intelligence Scale for Children.

Appendix K. Post-Hoc Sensitivity Analyses

K1: Multiple Regression with Picture Vocabulary and Demographic Factors Only

Predictor Variable	*B*(unstandardised)	Confidence Intervals (95%)		t	Sig.	Partial Eta Squared
Intercept	4.576	3.190,	5.961	6.473	<**.001**	.005
Sex	.653	.499,	.808	8.271	<**.001**	.008
Born in US (no)	-.480	-.932,	-.028	-2.082	**.037**	.001
NIH Picture Vocabulary	.001	-.004,	.006	.375	.708	.000
Household income	-.125	-.160,	-.090	-7.020	<**.001**	.006
Age	.093	-.032,	.217	1.456	.145	.000

Sample is unrelated participants including those with ASD/ADHD diagnoses

Abbreviations: NIH, National Institutes of Health

K2: Multiple Regression with WISC Matrix Reasoning and Demographic Factors Only in Males

Predictor Variable	*B*(unstandardised)	Confidence Interval (95%)		t	Sig.	Partial Eta Squared
Intercept	17.156	14.977,	19.334	15.436	<**.001**	.052
Born in US (no)	-1.223	-2.037	-.409	-2.946	**.003**	.002
WISC Matrix Reasoning	.005	-.042	.051	.193	.847	.000
Household income	-.245	-.303,	-.188	-8.361	<**.001**	.016
Age	-.051	-.266,	.163	-.470	.638	.000

Sample is male unrelated participants including those with ASD/ADHD diagnoses

Abbreviations: US, United States; WISC, Wechsler Intelligence Scale for Children

Appendix L. Reporting Checklist

STROBE Reporting Checklist for Cohort Studies

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	Pg 53 Pg57
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	Pg 58-60
Objectives	3	State specific objectives, including any prespecified hypotheses	Pg62-63
Methods			
Study design	4	Present key elements of study design early in the paper	Pg 64
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Pg 64
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up (*b*) For matched studies, give matching criteria and number of exposed and unexposed	Pg 65
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Pg 65-68
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Pg65-68
Bias	9	Describe any efforts to address potential sources of bias	Pg 69
Study size	10	Explain how the study size was arrived at	Pg 69
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Pg 63
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding (*b*) Describe any methods used to examine subgroups and interactions (*c*) Explain how missing data were addressed (*d*) If applicable, explain how loss to follow-up was addressed (*e*) Describe any sensitivity analyses	(a/b/e) Pg 68-70 (e) Pg 76
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed (b) Give reasons for non-participation at each stage (c) Consider use of a flow diagram	Pg 71/72 Pg 71/72
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (b) Indicate number of participants with missing data for each variable of interest (c) Summarise follow-up time (eg, average and total amount)	Pg 71/72
Outcome data	15*	Report numbers of outcome events or summary measures over time	Pg 71/72

Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included (*b*) Report category boundaries when continuous variables were categorized (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Pg 73-78
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Pg 82
Discussion			
Key results	18	Summarise key results with reference to study objectives	Pg 83-87
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Pg 91
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Pg 83-90
Generalisability	21	Discuss the generalisability (external validity) of the study results	Pg91
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Pg 92

Appendix M. Research Proposals

Appendix M1. Approved Major Research Proposal

https://osf.io/ywkef

Appendix M2. Approved Major Research Project Proposal (not completed)

https://osf.io/7rz5u

Appendix N. NIH Approval

Email from NDA Help Desk Evidencing Approval

Adding researchers to our ABCD project

ND NIMH Data Archive Help Desk <ndahelp@mail.nih.gov>
22/11/2023 16:45

Cc: Breda Cullen; Joey Ward; Joanna Loughrey (PGR); Liam Dorris

In replies all text above this line is added to the ticket

You are registered as a cc on this help desk request and are thus receiving email notifications on all updates to the request.
Reply to this email to add a comment to the request.

Lorraine Siochi (HELP DESK)
Nov 22, 2023, 11:45 AM EST

Good day,

We've successfully added access for Joanna. They need to accept the terms and conditions the next time they sign into NDA. Once they do so, they will be able to start accessing data.

Please let us know if you have any questions.

Thank you,

NDA Help Desk
NDAHelp@mail.nih.gov

To ensure NIMH Data Archive Help Desk emails reach your inbox, please add NDAHelp@mail.nih.gov to your email address book.