# Methods

**Participants**

180 children aged between 3.42 and 4.83 years of age (*M*=3.97, *SD*=0.25) participated (90 female) at baseline. Participants were recruited via local preschools, public advertising, social media and word of mouth using materials inviting families to take part in a project about children’s emotions when they start school. Given that their transition to school might be atypical, children with learning disabilities or special needs were not eligible to participate. Despite this, three parents reported at the baseline assessment that their child had any diagnosed learning difficulty, mental health problem or atypical development (one child had an Autism Spectrum Disorder diagnosis, one had a midline cleft pallet and the third was seeing a speech and language therapist). The data for these participants is retained but it is possible to identify which children these are based on the parent baseline questionnaire data variable ‘Diagnosis\_ND\_LD\_Other’. One participant withdrew from the study prior to the follow-up. All other participants were invited to participate in both follow-ups.

**Stages of the research**

1. **Baseline**

Baseline assessments were conducted when the child was in their preschool year. Parents completed baseline questionnaires during the baseline session. The eyetracking task, labTAB and cognitive data were collected during this session. Additional data using was collected using electroencephalography (EEG) and data from an anti-saccade task using eyetracking. Due to technical problems, these data are not year cleaned but will be added to the dataset when they are available. In the UK children begin school in the September when they are aged 4 years. Not all children start school on the same day and it is common for children to have some half days at school before they begin full-time. Participating children started school in either September 2017 or September 2018 and there is a ‘Cohort’ variable in each of the questionnaire data files identifying the year the child started school.

1. **Transition to school daily ratings**

State anxiety ratings were obtained via text messages sent daily to parents over the transition to school period. The text message read: Please reply with your rating for [Child’s name]’s anxiety today. Using a scale from 0 to 10, with 0 indicating no anxiety or worries at all and 10 indicating extremely anxious/worried. Parents were also asked to reply indicating whether their child had attended school that day and how much of a problem their child’s anxiety had been. Text messages were sent every evening at 7:30pm beginning two days before the child’s first day at school. Reminder texts were sent the following morning if no response was received. Children started school on day 3.

1. **Follow-up 1**

Follow-up 1 questionnaires were completed by parents and teachers online within the first half of the first term the child was in school. Parents were emailed approximately one week after their child started school asking them to complete a series of questionnaires online. They were sent a reminder email two weeks later if they had not yet completed the questionnaire. The online questionnaire was closed at the end of the half term (approximately 6 weeks after the children’s first day at school). Once the parent had provided the teacher’s contact details the child’s teacher was contacted and invited to take part in the research separately.

1. **Follow-up 2**

Follow-up 2 questionnaires were completed by parents and teachers within the first half of the second term the child was in school. Parents and teachers were emailed approximately one week after their child returned to school in term 2 asking them to complete a series of questionnaires online. They were sent a reminder email two weeks later if they had not yet completed the questionnaire. The online questionnaire was closed at the end of the half term (approximately 6 weeks after the children’s first day at school).

Note that no baseline teacher data is available as children had not yet started school.

**Questionnaire measures**

Data were collected on the following questionnaires. Table 1 shows which questionnaires were completed at which stage by parents and teachers.

**Children’s Behaviour Questionnaire (CBQ).** The effortful control scale of the Children’s Behaviour Questionnaire was completed by parents (Rothbart, Ahadi, Hershey & Fisher, 2001). Parents are asked to respond using a scale from 1 to 7 to indicate how true a description of the child’s reactions each item represents with reference to the past 6 months. The scale includes 26 items and provides an effortful control score where higher scores indicate more effortful control. Full details and instructions for requesting to use the measure can be found here: <https://research.bowdoin.edu/rothbart-temperament-questionnaires/instrument-descriptions/the-childrens-behavior-questionnaire/>

**Preschool Anxiety Scale (PAS).** The Preschool Anxiety Scale (PAS; Spence, Rapee; McDonald & Ingram, 2001) was used to obtain parent and teacher report of child anxiety symptoms. The parent version is designed for parents of children aged 2.5 to 6.5 years and includes 28 items that provide an overall measure of child anxiety as well as five subscales reflecting symptoms of a range of child anxiety disorders. The measure has good construct validity, satisfactory internal consistency and good cross-informant and test-retest reliability (Spence et al., 2001). The teacher version includes 22 items also scored on a 5-point likert scale. Five subscale scores are calculated and one total score, to parallel the parent measure. Full details and the measure can be found at www.scaswebsite.com.

**MacArthur Health and Behavior Questionnaire (HBQ).** Parents and teachers completed the Mental Health, Social Functioning and School functioning sections of the HBQ (Boyce et al., 2002; Essex et al., 2002). The measure is designed to collect data from adults about children aged 4 to 8 years. It yields a number of subscales including anxiety, depression, separation anxiety, oppositional defiant, conduct problems, prosocial behavior, hostility, aggression, inattention, hyperactivity, peer acceptance, bullying, social inhibition and asocial behaviour. These can be combined to give overall internalizing and externalizing scores. Full details and the measure can be found at: <https://macarthurhbq.wordpress.com/>

**Behavioural Inhibition Questionnaire (BIQ).** The BIQ (Bishop, Spence & McDonald, 2003) was used to obtain parent report of child behavioural inhibition (BI). The BIQ contains 30 items relevant to child BI, with some reverse scored items. Parents are asked to indicate using a scale from 1 (Hardly ever) to 7 (Almost always) how often each behavior occurs for their child. The measure has good psychometric properties, with adequate internal consistency, moderate stability of time and strong construct validity (Bishop et al., 2003). Full details and the measure can be found at [www.scaswebsite.com](http://www.scaswebsite.com).

**Child Autism Spectrum Quotient-10 (AQ).** The 10-item Child AQ (Allison, Auyeung & Baron-Cohen, 2012) was completed. Parents were asked to indicate the extent to which they agreed that each item characterized their child. The scale gives an overall score with higher scores indicating an increased likelihood that the child may need to be referred for an assessment for Autism Spectrum Disorder.

**Social Motivation Questionnaire (SMQ).** A measure of child social motivation was created by adjusting the Anticipatory and Consummatory Interpersonal Pleasure Scale (ACIPS; Gooding & Pflum, 2014a; 2014b), which is designed to measure hedonic capacity for social and interpersonal pleasure. The ACIPS is designed for adults so the questions were adjusted by the grant holder to be suitable for children. The final questionnaire contained 15 items which were scored on a scale from 1 to 6 to indicate how true each item is for the child.

**Empathy Questionnaire (EQ). T**he Empathy Questionnaire (EmQue; Rieffe et al., 2010) is a 20 item questionnaire designed for parents to report on children up to six years of age. Each item is a statement concerning the child's empathy-related behaviour over the past two months. Parents can respond with one of three options (no/sometimes/often). The EQ consists of three different sub-scales: i) Emotion contagion (6 items); ii) Attention to others' emotions (7 items); and iii) Prosocial responses to others' emotions (6 items).

**Children’s Responses to Uncertainty and Low Environmental Structure (RULES).** The RULES (Sanchez et al., 2017) is a 17-item parent report measure of intolerance of uncertainty (IU) in young children. It is designed to measure IU in children aged 3-10 years. Parents are asked to indicate how well each statement characterizes their child on a scale from 1 to 5. The scale gives a total score where higher scores indicate greater intolerance of uncertainty.

**State Trait Anxiety Inventory (STAI).** The STAI (Speilberger et al., 1983) is a commonly used measure of state and trait anxiety. Parents only completed the 20 trait anxiety questions. Questions are answered on a 4 point likert scale from “Almost never” to “Almost always”. Higher scores indicate higher levels of trait anxiety.

**Social Phobia Inventory (SPIN).** The SPIN (Connor et al., 2000) assesses social anxiety in adults. It evaluates fear, avoidance and physiological symptoms. The measure includes 17 items, which are scored from 1 to 5 on a 5 point likert scale. Higher scores indicate greater social anxiety.

**School Anxiety Scale (SAS).** The SAS (Lyneham et al., 2008) is designed to collect information on children’s anxiety at school from teachers. The questionnaire contains 16 items and teachers respond to indicate how often each statement described the child over the previous three months, using a 4 point Likert scale.

Table 1: Questionnaires completed at each stage by parents and teachers.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Parent baseline** | **Parent follow-up 1** | **Parent follow-up 2** | **Teacher follow-up 1** | **Teacher follow-up 2** |
| CBQ | Y | N | N | N | N |
| PAS | Y | Y | Y | Y | Y |
| HBQ | Y | Y | Y | Y1 | Y[[1]](#footnote-1) |
| BIQ | Y | N | N | N | N |
| AQ | Y | N | N | N | N |
| SMQ | Y | N | N | N | N |
| EQ | Y | N | Y | N | N |
| RULES | Y | Y | Y | N | N |
| STAI | Y | Y | Y | N | N |
| SPIN | Y | Y | Y | N | N |
| SAS | N | N | N | Y | Y |

**Variable names and labels in questionnaire data files.**

Across the questionnaire files, variable names begin with ‘a’ to indicate baseline data, ‘b’ to indicate follow-up 1 data and ‘c’ to indicate follow-up 2 data. The letters that follow indicate the questionnaire acronym (see brackets after each questionnaire name above and in Table 1) and the number indicates the item number on the questionnaire. For example, bPAS14 is follow-up data for the Preschool Anxiety Scale, item 14. Teacher and parent data uses the same format with the exception that the teacher data includes the letter ‘t’ after the initial letter (e.g. btPAS14). Across all data files, variable labels are used to explain what the variable is. For questionnaire data this will be the exact wording of the item from the questionnaire, with the exception of the total and scale scores. Syntax files are provided to demonstrate how total and scale scores were computed.

**Additional measures completed at baseline assessment**

The following additional measures were completed at the baseline assessment. Data for each is provided in a separate data file.

**Observation of BI.** The Laboratory Temperament Assessment Battery (Lab-TAB) is a standardised observational measurement of temperament designed for use with 3-5 year old children (Gagne, Van Hulle, Aksan, Essex, & Goldsmith, 2011). The present study used two of the four available fear episodes most relevant to BI (Risk Room and Stranger Approach). During phase one of the risk room, the child was free to explore and play in an unfamiliar room with five novel/ambiguous toys (e.g. balance beam; tunnel; black box; wolf mask; steps and mattress). During phase two a research assistant asked the child to play with each toy in turn. Coding included time to touch objects, time spent playing with objects, weary facial affect, tentativeness or play and approaching/referencing parent. During the stranger approach episode the child was approached by a ‘stranger’ as they waited alone in a room. The Lab-TAB manual specifies a male stranger, but this was not feasible so instead an unfamiliar female researcher acted as the stranger and wore a boiler suit with a hood to appear slightly more threatening. Full details of the procedure can be found in the Lab-TAB manual (Goldsmith, Reilly, Lemery, Longley, & Prescott, 1999).

Episodes were videotaped and coded according to the Lab-TAB manual. Scores for each coding criteria were reverse coded as necessary so that higher scores indicated more inhibited responses. These scores were then converted into z-scores individually and then averaged into a single BI score in line with previous research (Gagne et al., 2011; Goldsmith, Reilly, Lemery, Longley, & Prescott, 1999). Larger, positive values indicated higher levels of BI. To check reliability a secondary coder coded 24% of the Lab-TAB assessments. Interrater reliability was good to excellent (single measure ICC=.95; 96%CI =.91 to .98). Both coders were blind to the child’s anxiety scores and bias scores.

**NIH Toolbox Early Childhood Cognition Batter v1.8.** Children’s cognitive functioning was assessed using the NIH Toolbox Early Childhood Cognition Battery v1.8, which is validated for use with 3-6 year old children. It produces scores for individual subtests as well as an Early Childhood Composite Score. Four subtests were completed: Vocabulary, Flanker Inhibitory Test, Dimensional Change Card Sort and Picture Memory subtests. Full details of both measures can be found in Zelazo et al. (2013). The toolbox provides uncorrected standard scores and age-corrected standard scores for the subtests and overall composite score, all of which are included in the data file. Both age-corrected and uncorrected standard scores have a mean of 100 and a standard deviation of 15. The age-corrected scores are adjusted relative to age norms and the uncorrected scores are adjusted relative to the entire population of norms, without reference to age.

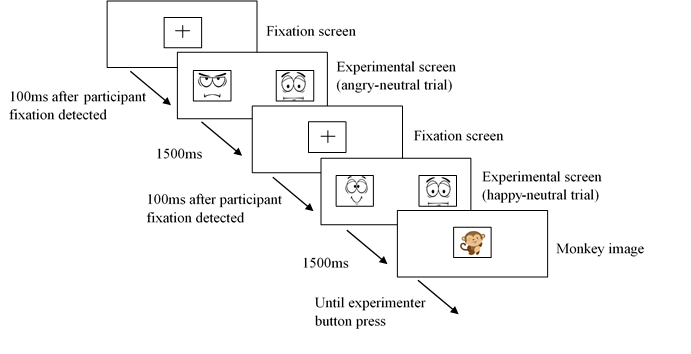
**Eye-tracking attention bias task.**Eye movements were measured using a remote desk-mounted Tobii T300 eye-tracker. The task was programmed in e-prime.

Colour photographs of 16 Caucasian child actors (8 male and 8 female) were sourced from the Child Affective Facial Expression set (CAFÉ; LoBue & Thrasher, 2015). Each actor posed for neutral, angry and happy facial affects. Happy faces were included as well as angry faces to ensure any observed effects could be attributed to the threatening, angry faces specifically as opposed to emotional faces in general. The faces chosen had the highest available accuracy ratings. The CAFE images are aligned based on each actor’s individual eye alignment. The photographs were manipulated so that for each actor, the three emotion images were matched for RGB values as well as mean and standard deviation luminance. Photographs were resized to 178x187 at 60 ppi. The face stimuli were positioned to centre at 761x540 and 1159x540. The viewing distance was approximately 65cm, giving a viewing angle of 4.4 degrees to the centre of the screen, and 9.4 degrees between the centre points of the face stimuli.

The task began with a 9-point calibration. Calibration was accepted as adequate if the error shown was within 1 degree for all 9 points. If this was not achieved, calibration was conducted again and the position of the eyetracker was adjusted until acceptable calibration was achieved.

Each trial began with a fixation screen showing a central cross surrounded by a 2” square purple outline to highlight the fixation cross. Gaze contingency was used such that the experimental screen began only after the participant had fixated within the square for a minimum of 100ms. The fixation screen was then replaced by the experimental screen for 1500ms. The experimental screen consisted of an emotion (happy/angry) and neutral face displayed side by side of the same child (positions counterbalanced). In total there were 64 experimental trials (32 Happy-Neutral; 32 Angry-Neutral).

To make the task engaging for children, on 14 additional trials the fixation screen was followed by a ‘monkey’ screen. Here a cartoon monkey was presented centrally and participants were asked to find the monkey on a tick chart. Fifteen cartoon monkey images were used, each showing the monkey in a different position (hanging from a branch, sitting arms crossed etc.). The ordering of experimental and monkey screens was randomized, with the exception of the final screen, which was always a monkey screen with the message ‘Well done!’. Monkey screens remained shown until the experimenter pressed a resume button. The monkey screens were included solely to engage the children; no data from these screens is analysed. See Figure 1 for an illustration of the eye-tracking task.



*Figure 1.* Illustration of the eye-tracking task. Note the face stimuli in this illustration were not the actual stimuli used, but are representative of the different emotions.

Participants were told that they were going to play a game using a computer that knew where they were looking. They were shown a sheet with 15 cartoon monkeys and were told they needed to look out for each of these monkeys on the screen. Participants were told that lots of faces would also appear on the screen but their job was to look out for the monkeys. When a monkey appeared the task was paused to allow them time to find which monkey they had seen on the score sheet and cross it off. It was hoped the use of a score sheet would encourage engagement and minimise off-screen eye movements. One of the researchers sat with the participant throughout the task to encourage engagement, provide support with the score sheet and monitor their well-being.

Eye movement data was continuously recorded throughout the entire period of stimulus presentation (1500ms) at a sampling rate of 300Hz. Three Areas of Interest (AOIs) were defined based on the location of the two face images and the square surrounding the fixation cross. The raw eye-tracking data were processed in R version 3.5.3 using the EyetrackingR package (Dink & Ferguson, 2018) and initial checks were conducted to ensure data quality. Fixation plots were visually inspected against the position of stimuli on the display screen and five additional participants were excluded given evidence of poor calibration. Individual trials were removed if there was greater than 40% trackloss for that trial. On average, the included sample had 59.3 valid trials and <8% trackloss.

***Calculation of bias scores presented in data file.***Orientation bias for angry faces was calculated as the percentage of angry-neutral trials where participants looked to a face, in which gaze was recorded within the angry face AOI first. The equivalent was also calculated for happy faces. To calculate dwell bias for angry faces, the number of samples recorded in the angry face AOI and neutral face AOI during angry-neutral trials were calculated as a proportion of the total number of samples recorded during each trial and the resulting proportion of looking to the neutral face was subtracted from proportion of looking to the angry face. The equivalent was also calculated for happy dwell bias using happy-neutral trials.

**Ethics**

All study procedures were approved by the University of Reading ethics committee and were performed in accordance with the ethical standards outlined by the Declaration of Helsinki. Informed consent and assent were obtained from parents, teachers and participants after being provided with information about the project in a format suited to their cognitive and language abilities.Parents consented to take part in the follow-up stages of the research during the baseline session. Copies of consent forms are provided.

**Procedure**

Families were invited to a lab session at the University of Reading. During the session participants completed the tasks reported above along with a task using electroencephalography (EEG) and an anti-saccade eyetracking task. Data for these tasks will be uploaded to the reshare repository once it is cleaned. The session began with the Lab-TAB, which was followed by the NIH Toolbox vocabulary test then the eyetracking task and finally the remaining Toolbox subtests. Participants were given breaks between tasks and were rewarded with a sticker after the completion of each task. At the end of the lab session participants were thanked for their participation and received small monetary value.

**Missing data**

**Questionnaires.** Data are missing from questionnaires only where a parent or teacher has chosen not to complete an item or the questionnaire entirely; usually because they did not participate in the follow-up. Note that teacher data is missing for approximately 1/3 of participants because the teacher chose not to participate in the research.

**LabTAB.** Data are missing for 16 participants on the LabTAB due to problems during administration (e.g. camera malfunction, child choosing to withdraw from task, experimenter error).

**Cognitive Assessment.**Missing data occur in the cognitive assessment when a child failed to reach scoring threshold on the subtests.

**Eyetracking.**Four participants did not start the task due to problems calibrating and three were excluded because they did not complete the task. Five additional participants were excluded because of poor calibration (see above).

**Text message anxiety ratings.**Text message anxiety ratings were received for the all 179 participants who remained in the study when the children started schools; 150 responded on all 14 days. The mean number of days rating given was 13.56 (SD = 1.19; Min = 7 days).

**References**

Allison, C., Auyeung, B., & Baron-Cohen, S. (2012). Toward brief “red flags” for autism screening: the short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(2), 202-212.

Bishop, G., Spence, S. H., & McDonald, C. (2003). Can parents and teachers provide a reliable and valid report of behavioral inhibition?. *Child Development*, *74*(6), 1899-1917.

Boyce, W.T., Essex, M.J., Woodward, H.R., Measelle, J.R., Ablow, J.C., Kupfer, D.J., & The MacArthur Assessment Battery Working Group. (2002). The confluence of mental, physical, social, and academic difficulties in middle childhood. I: Exploring the “headwaters” of early life morbidities. *Journal of the American Academy of Child and Adolescent Psychiatry, 41,* 580-587.

Connor, K. M., Davidson, J. R., Churchill, L. E., Sherwood, A., Weisler, R. H., & Foa, E. (2000). Psychometric properties of the Social Phobia Inventory (SPIN): New self-rating scale. *The British Journal of Psychiatry*, *176*(4), 379-386.

Essex, M.J., Boyce, W.T., Goldstein, L.H., Armstrong, J.M., Kraemer, H.C., Kupfer, D.J., & The MacArthur Assessment Battery Working Group. (2002). The confluence of mental, physical, social, and academic difficulties in middle childhood. II: Developing the MacArthur Health and Behavior Questionnaire*. Journal of the American Academy of Child and Adolescent Psychiatry, 41,* 588-603.

Gagne, J. R., Van Hulle, C. A., Aksan, N., Essex, M. J., & Goldsmith, H. H. (2011). Deriving childhood temperament measures from emotion-eliciting behavioral episodes: Scale construction and initial validation. Psychological Assessment, 23(2), 337.

Goldsmith, H., Reilly, J., Lemery, K.S., Lonley, S., & Prescott, A. (1999). The Laboratory Temperament Assessment Battery: Preschool Version. Version 0.5.

Gooding, D. C., & Pflum, M. J. (2014a). The assessment of interpersonal pleasure: introduction of the Anticipatory and Consummatory Interpersonal Pleasure Scale (ACIPS) and preliminary findings. *Psychiatry research*, *215*(1), 237-243.

Gooding, D. C., & Pflum, M. J. (2014b). Further validation of the ACIPS as a measure of social hedonic response. *Psychiatry research*, *215*(3), 771-777.

LoBue, V., & Thrasher, C. (2015). The Child Affective Facial Expression (CAFE) set: validity and reliability from untrained adults. *Frontiers in psychology*, *5*, 1532.

Lyneham, H. J., Street, A. K., Abbott, M. J., & Rapee, R. M. (2008). Psychometric properties of the school anxiety scale—Teacher report (SAS-TR). *Journal of Anxiety Disorders*, *22*(2), 292-300.

Rieffe, C., Ketelaar, L., & Wiefferink, C.H. (2010). Assessing empathy in young children; construction and validation of an empathy questionnaire (EmQue*). Personality and Individual Differences, 49,* 362–367.

Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at 3-7 years: The Children’s Behavior Questionnaire. *Child Development, 72,* 1394-1408.

Sanchez, A.L. Cornacchio, D., Chou, T., Leyfer, O., Coxe, S., Pincus, D., & Comer, J.S. (2017). Development of a scale to evaluate young children’s resopnses to uncertainty and low environmental structure. *Journal of Anxiety Disorders, 45,* 17-23.

Spence, S. H., Rapee, R., McDonald, C., & Ingram, M. (2001). The structure of anxiety symptoms among preschoolers. *Behaviour Research and Therapy, 39(11),* 1293-1316.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Zelazo, P. D., & Bauer, P. J. (Eds.). (2013). *National Institutes of Health Toolbox cognition battery (NIH Toolbox CB): Validation for children between 3 and 15 years*. Wiley.

1. There was a mistake when the questions for this were uploaded which means that there are a number of additional questions that were from the parent HBQ. These have ‘add’ added to the end of the variable name and are not used for the scale scores. In addition, item 87 (Complains of stomach aches and headaches) and item 129 (Complains about not feeling well) from the teacher HBQ were not included as exactly written. In the scale scores these are replaced with ‘Has aches and pains’ and ‘Has nausea, feels sick’ respectively. [↑](#footnote-ref-1)