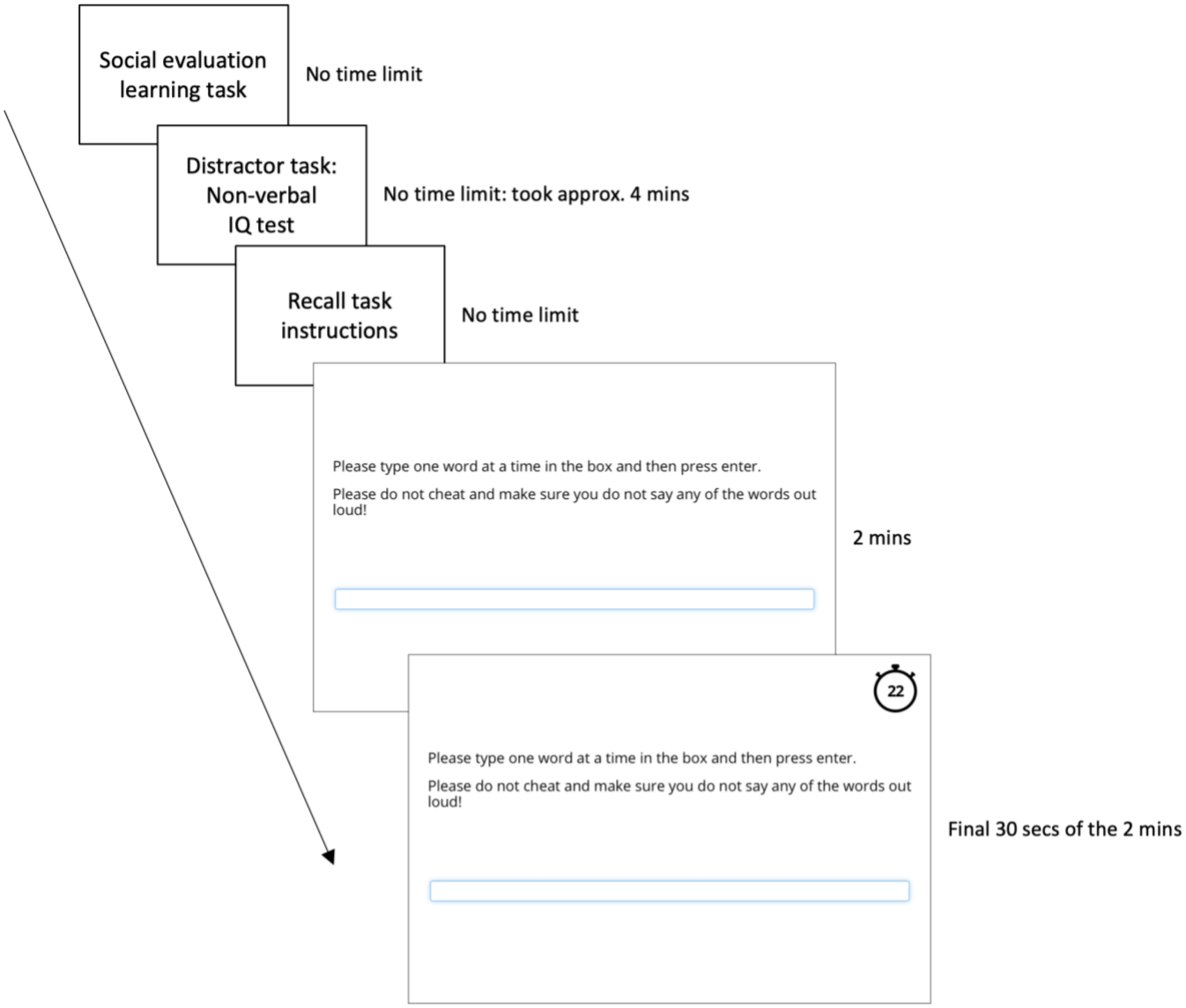
Supplementary Materials

Methods

Surprise recall task

As shown in Supplementary Figure 1, participants first did a self-paced social evaluation learning task. They then completed the non-verbal IQ test, which served as a distractor task between learning the personality descriptors and the surprise recall test. On average, the non-verbal IQ test took 4mins. Participants then read the detailed instructions for the surprise recall task, which informed them that the time limit would start when they continued to the next page. Once they clicked through to the next page, participants had 2mins to type as many words as they could recall from the social evaluation learning task. A timer appeared in the top right of the screen for the final 30secs of this recall test, which counted down from 30 to 0 as they reached the time limit.



Supplementary Figure 1. Study procedure with task timings. The final two screens show the surprise recall task as seen by participants.

Depressive symptoms

To check that gender differences in recall were not explained by depressive symptoms, we added SMFQ score to the negative binomial mixed models testing the effect of age group and gender on recall. We first tested a three-way interaction between gender, condition and valence with hits as the dependent variable, and also report the two-way interactions between these variables. We then tested a four-way interaction between age group, gender, condition and valence with hits as the dependent variable, and report the two-way and three-way interactions between these factors. As our aim was to compare the influence of gender in each age group, we only report the lower level (two-way and three-way) interactions which include gender. Where there was evidence of an interaction, we examined associations with hits separately for each subgroup. All models were adjusted for depressive symptoms, condition, valence, continuous age within each group, school, testing group size, non-verbal IQ score, and positive and negative false alarms.

Additional confounders

After completing the consent form, participants' parents/carers were asked to complete a questionnaire via Gorilla (www.gorilla.sc). Parents/carers reported participants' ethnicity, first language, whether they had been diagnosed with a mental health problem, diagnoses (if applicable), use of mental health services, special educational needs and disabilities (including diagnoses of dyslexia and autism spectrum disorders (ASD)), current use of antidepressants and psychological therapy for depression, parental education, and parental mental health (specifically experiences of depression, anxiety, or stress).

Ethnicity was grouped into 5 categories based on the most common ethnicities – White, Mixed, Asian/Asian British, Black/Black British, and Other. First language was categorised as a binary variable to indicate whether the participant's first language was English (0 = no, 1 = yes). As an indicator of socioeconomic status, we took the highest reported parental education for each participant (across both mother and father). Parental education was then split into low (highest qualification GCSE or lower) and high (A Levels or higher). As our tasks involved skills that are known to be affected in dyslexia and ASD, we created binary variables to indicated whether participants had been diagnosed with dyslexia or ASD, as reported by parents/carers.

We first explored the distribution of these confounders by age group and tested whether they were associated with task performance and depressive symptoms. We then repeated primary analyses for the subsample whose parents/carers completed the questionnaire. Results are presented before and after controlling for the additional confounders (ethnicity, English as a first language, dyslexia, ASD, parental education, maternal depression, paternal depression) in this subsample. We used negative binomial mixed models testing the effects of condition, valence, age group, and gender on hits and the linear regression model testing associations between hits and depressive symptoms.

Pubertal stage

We intended to adjust for pubertal stage as it is strongly associated with depressive symptoms (Angold, Costello, & Worthman, 1998). Pubertal hormones and age have functionally dissociable effects on neural activity during social emotion processing (Goddings, Burnett Heyes, Bird, Viner, & Blakemore, 2012). There is also evidence that puberty is more strongly associated with performance on a self-referential encoding task than age (Ke, Wu, Willner, Brown, & Crowley, 2018).

The Pubertal Development Scale (PDS) is a self-report measure of pubertal status (Petersen, Crockett, Richards, & Boxer, 1988). It consists of three questions about the development of body hair, the occurrence of a growth spurt, and changes in complexion plus two sex-specific items. Girls were asked about breast development and the onset of menstruation and boys were asked about changes in voice and growth of facial hair. Responses are coded on 4-point scales (1 = no development and 4 = completed development) except for the question about onset of menarche, which was rated on a 3-point scale (1 = no and 3 = yes definitely). Following classroom data collection, participants were emailed a link via Gorilla to complete the PDS at home (because of the sensitive content).

Pubertal development is traditionally classified into five Tanner stages - prepubertal, early pubertal, mid-pubertal, late pubertal, and post-pubertal (Carskadon & Acebo, 1993; Crockett, 1988; Norris & Richter, 2008). However, given the unbalanced number of participants across the five groups, particularly the small number of prepubertal and early pubertal participants (n=14), we divided participants into two groups: early/mid puberty (Tanner stages 1-3) and late/post puberty (stages 4-5). Girls in the early group were pre-menarche and girls in the late group were post-menarche. Boys in the early group had low individual ratings on growth of body hair, voice change, and growth of facial hair growth compared to boys in the late group.

We first explored whether pubertal stage was associated with task performance and depressive symptoms. We then repeated all primary analyses in this subsample, with and without adjusting for pubertal stage. Negative binomial mixed models would not converge when pubertal stage was included (likely due to overspecification in the reduced sample size) so Poisson mixed models were used.

Results

Depressive symptoms

Adding depressive symptoms to the negative binomial mixed models testing the associations between condition, valence, gender, age group and hits did not alter the evidence for any of the associations (Supplementary Table 2).

Additional confounders

Demographic information was only available for a subsample of participants (n=340, 59%; Supplementary Table 1). In this subsample, 78% were of white ethnicity and 89% had English as their first language. According to parents/carers, 6% of the subsample had a mental health problem, with 1% reporting depression. Use of mental health services by their child was reported by 8% of parents/carers, and 1% of participants were reported to be currently receiving psychological therapy for depression. Only 1 participant (0%) was currently using antidepressants according to parental report. In total, 14% of parents/carers reported that their child had special educational needs and disabilities, with 7% reporting dyslexia and 1% reporting ASD. Parent education was high for 88% of the subsample. In terms of mental health, 44% of mothers and 34% of fathers had experienced depression, anxiety or stress; 16% of mothers and 9% of fathers had depression specifically.

The two age groups were similar except that mid-adolescents had higher parent-reported rates of mental health problems, use of mental health services, and receipt of psychological therapy (Supplementary Table 1).

Associations with recall

In separate negative binomial mixed models, there was no evidence that ethnicity (p=0.33), English as first language (p=0.33), parental education (p=0.64), maternal depression (p=0.93), or ASD (p=0.12) were associated with hits. However, there was evidence that paternal depression (hits ratio = 0.82, 95% CI = 0.69 to 0.97, p=.02) and parent-reported diagnoses of dyslexia (hits ratio = 0.65, 95% CI = 0.53 to 0.79, p<0.001) were associated with poorer recall (fewer hits).

We first repeated the main analyses in the subsample for whom additional confounders were not missing. In this subsample there was only evidence that valence and gender were associated with total hits (Supplementary Table 3). Participants made 10% (95% CI=1.03 to 1.19, p<0.001 adjusted for confounders) more negative than positive hits. There was evidence that females made 18% (95% CI=1.08 to 1.30, p<0.001 adjusted for confounders) more hits than males. In contrast to analyses in the whole sample, there was no evidence that condition or age group were associated with total hits in this subsample. Adjusting for additional confounders did not alter the evidence for any of these associations (Supplementary Table 3). Also, in this subsample, there was no evidence for any interactions between age group, gender, condition, or valence on total hits, before and after adjusting for additional confounders (Supplementary Table 4).

Associations with depressive symptoms

In separate linear regression models, there was no evidence that English as first language (p=0.24), parental education (p=0.60), maternal depression (p=0.70), paternal depression (p=0.14), dyslexia (p=0.22), or ASD (p=0.62) were associated with adolescent-reported depressive symptoms (SMFQ score). There was evidence that ethnicity was associated with depressive symptoms (p=0.02). Compared to white participants (M=7.34, SD=5.39), mixed (M=9.11, SD=7.69), Asian/Asian British (M=10.21, SD=8.75), and other ethnicities (M=10.24, SD=5.30) had higher depressive symptoms. The Black/Black British group reported lower depressive symptoms (M=5.83, SD=3.32).

In this subsample, there was only evidence that self-referential negative hits were associated with depressive symptoms (coef = 0.65, 95% CI = 0.20 to 1.10, p=0.005 adjusted for additional confounders). Before and after adjusting for additional confounders, there was no evidence that self-referential positive hits were associated with depressive symptoms (Supplementary Table 5). Consistent with the main analysis, there was no evidence that other-referential positive or negative hits were associated with depressive symptoms.

For all analyses in this subsample, before and after adjusting for additional confounders, the effect estimates were similar to the coefficients, and within the confidence intervals, from the primary analyses with the whole sample. The lack of evidence could be due to the reduced sample size or selection bias in those whose parents/carers completed additional questions.

Pubertal stage

The Pubertal Development Scale (PDS; Petersen et al., 1988) was fully completed by a small proportion of participants (n=117, 20% of total sample). Of the subsample that completed this measure, 61 (52%) were in early pubertal stages (pre-pubertal, early pubertal, and mid-pubertal) and 56 (48%) were in late pubertal stages (late pubertal and post-pubertal). Mid-adolescents were generally in later stages of puberty than young adolescents (Supplementary Table 1).

Associations with recall

In an unadjusted Poisson mixed model, there was no evidence that pubertal stage was associated with total hits (hits ratio = 1.14, 95% CI 0.99 to 1.32, p=0.07). However, given that this evidence for an association between pubertal stage and hits narrowly missed statistical significance (p=0.05), we repeated the main analyses adjusting for pubertal stage.

In the subsample of participants who completed the PDS, there was only evidence that valence and gender were associated with total hits (Supplementary Table 3). Participants made 13% (95% CI=1.02 to 1.26, p=0.02 adjusted for confounders) more negative than positive hits. Females made 18% (95% CI=1.08 to 1.30, p<0.001 adjusted for confounders) more hits than males. In contrast to analyses in the whole sample, there was no evidence that condition or age group were associated with total hits. Adjusting for pubertal stage did not alter the evidence for any of these associations (Supplementary Table 3). In this subsample, there was also no evidence for any interactions between age group, gender, condition, or valence on total hits, before and after adjusting for pubertal stage (Supplementary Table 4).

Associations with depressive symptoms

There was evidence that pubertal stage was positively associated with depressive symptoms in a linear regression model (coef = 2.30, 95% CI 0.23 to 4.38, p=0.03). Adolescents in late, compared to early, pubertal stages scored 2.30 (95% CI 0.23 to 4.38, p=0.03) points higher on the SMFQ. In this subsample, there was no evidence that any type of hits was associated with depressive symptoms, before and after adjusting for pubertal stage (Supplementary Table 5).

As with the previous sensitivity analyses, in the subsample who reported pubertal stage, coefficients were similar to (and within the confidence intervals from) findings in the primary analysis. The lack of evidence could be due to the reduced sample size or selection bias in participants who completed the PDS.

References

Angold, A., Costello, E. J., & Worthman, C. M. (1998). Puberty and depression: The roles of age, pubertal status and pubertal timing. *Psychological Medicine*, *28*(1), 51–61. https://doi.org/10.1017/S003329179700593X

Carskadon, M. A., & Acebo, C. (1993). A Self-Administered Rating Scale for Pubertal Development. *Journal of Adolescent Health*, *14*, 190–195.

Crockett, L. J. (1988). *Pubertal Development Scale: Pubertal Categories*. Unpublished manuscript, Pennsylvania State University, Department of Human Development and Family Studies,

University Park.
Goddings, A. L., Burnett Heyes, S., Bird, G., Viner, R. M., & Blakemore, S. J. (2012). The relationship between puberty and social emotion processing. *Developmental Science*, *15*(6), 801–811. https://doi.org/10.1111/j.1467-7687.2012.01174.x
Ke, T., Wu, J., Willner, C. J., Brown, Z., & Crowley, M. J. (2018). Adolescent positive self, negative self: associated but dissociable? *Journal of Child and Adolescent Mental Health*, *30*(3), 203–211. https://doi.org/10.2989/17280583.2018.1552590
Norris, S. A., & Richter, L. M. (2008). Are There Short Cuts to Pubertal Assessments? Self-Reported and Assessed Group Differences in Pubertal Development in African Adolescents. *Journal of Adolescent Health*, *42*(3), 259–265. https://doi.org/10.1016/j.jadohealth.2007.08.009
Petersen, A. C., Crockett, L., Richards, M., & Boxer, A. (1988). A self-report measure of pubertal status: Reliability, validity, and initial norms. *Journal of Youth and Adolescence*, *17*(2), 117–133. https://doi.org/10.1007/BF01537962

Supplementary Table 1. *Characteristics of the subsample with parental report according to age group, n (%) reported.*

	Young adolescents	Mid-adolescents	Overall
Ethnicity			
White	150 (79%)	115 (77%)	265 (78%)
Mixed	9 (5%)	16 (11%)	25 (7%)
Asian/Asian British	5 (3%)	2 (1%)	7 (2%)
Black/Black British	7 (4%)	3 (2%)	10 (3%)
Other	18 (10%)	13 (9%)	31 (9%)
Total n	189	149	338
English as first language	168 (90%)	130 (88%)	298 (89%)
Total n	186	148	334
Mental health problem	6 (3%)	13 (9%)	19 (6%)
Total n	190	149	339
Used mental health services	12 (6%)	16 (11%)	28 (8%)
Total n	190	149	339
Receiving psychological therapy	0 (0%)	5 (3%)	5 (1%)
Total n	190	150	340
Special educational needs and disabilities	24 (13%)	23 (16%)	47 (14%)
Total n	188	148	336
High parental education	159 (88%)	122 (88%)	281 (88%)
Total n	180	139	319
Maternal depression anxiety or stress	78 (44%)	64 (45%)	142 (44%)
Total n	177	143	320
Paternal depression anxiety or stress	54 (34%)	42 (34%)	96 (34%)
Total n	161	123	284
Pubertal stage			
Early	48 (84%)	13 (22%)	61 (52%)
Late	9 (16%)	47 (78%)	56 (48%)
Total n	57	60	117

Note. Mental health problem denotes whether participant had ever been diagnosed with any mental health problem (anxiety, conduct disorder, depression, eating disorder, generalized anxiety disorder, panic disorder, PTSD, social anxiety, substance misuse, other disorder). Highest reported parental education for each participant (across both mother and father) was split into low (highest qualification GCSE or lower) and high (A Levels or higher). Pubertal stage was split into early (pre-pubertal, early pubertal, and mid-pubertal) and late (late pubertal and post-pubertal).

Supplementary Table 2. *Negative binomial mixed models testing the effect of age group, gender, condition (whether words were learnt in relation to the self or another person) and valence (whether words were positive or negative) on the total number of hits, adjusted for depressive symptoms.*

	n	Interaction	Subgroup analysis	
		p value	Hits ratio	95% CI
Gender x condition	**566**	**0.04**		
Males: condition	286		1.14	1.05 to 1.23
Females: condition	280		1.02	0.94 to 1.09
Gender x valence	**566**	**0.99**		
Males: valence	286		1.14	1.05 to 1.23
Females: valence	280		1.14	1.06 to 1.22
Gender x condition x valence	**566**	**0.87**		
Males self-referential: valence	286		1.11	1.00 to 1.24
Males other-referential: valence	286		1.17	1.04 to 1.31
Females self-referential: valence	280		1.10	0.99 to 1.22
Females other-referential: valence	280		1.18	1.06 to 1.31
Age group x gender	**566**	**0.13**		
Young adolescents: gender	304		1.04	0.93 to 1.17
Mid-adolescents: gender	262		1.26	1.12 to 1.41
Age group x gender x condition	**566**	**0.05**		
Young adolescent males: condition	158		1.18	1.05 to 1.31
Young adolescent females: condition	146		0.93	0.83 to 1.04
Mid-adolescent males: condition	128		1.10	0.99 to 1.24
Mid-adolescent females: condition	134		1.09	0.99 to 1.20
Age group x gender x valence	**566**	**0.43**		
Young adolescent males: valence	158		1.14	1.02 to 1.27
Young adolescent females: valence	146		1.19	1.06 to 1.33
Mid-adolescent males: valence	128		1.14	1.02 to 1.28
Mid-adolescent females: valence	134		1.10	0.99 to 1.21
Age group x gender x condition x valence	**566**	**0.42**		

Note. All models adjusted for depressive symptoms, condition, valence, gender, age group, continuous age within each group, school, testing group size, non-verbal IQ score, and positive and negative false alarms. For gender, males were the reference group. For condition, other-referential hits were the reference group. For valence, positive hits were the reference group.

Supplementary Table 3. *Negative binomial and Poisson mixed models testing the effect of age group, gender, condition (self-/other-referential) and valence (positive/negative) on total hits, in subsamples with additional confounders or pubertal stage reported.*

	Subsample with data on additional confounders N=275				**Subsample with data on pubertal stage** N=117			
	Model 1: Adjusted		**Model 2: Additionally adjusted**		**Model 1: Adjusted**		**Model 2: Additionally Adjusted**	
	Hits ratio (95% CI)	p value	Hits ratio (95% CI)	p value	Hits ratio (95% CI)	p value	Hits ratio (95% CI)	p value
Condition	1.06 (0.99 to 1.14)	0.10	1.06 (0.99 to 1.14)	0.10	1.02 (0.91 to 1.13)	0.76	1.02 (0.91 to 1.13)	0.76
Valence	1.10 (1.03 to 1.19)	0.01	1.10 (1.03 to 1.19)	0.01	1.13 (1.02 to 1.26)	0.02	1.13 (1.02 to 1.26)	0.02
Age group	1.31 (0.96 to 1.79)	0.09	1.34 (0.99 to 1.81)	0.06	1.09 (0.72 to 1.64)	0.69	1.17 (0.76 to 1.79)	0.47
Gender	1.18 (1.03 to 1.19)	<0.001	1.18 (1.08 to 1.30)	<0.001	1.18 (1.04 to 1.34)	0.01	1.22 (1.06 to 1.40)	0.004

Note. All models were adjusted for condition, valence, gender, age group, continuous age within each group, school, testing group size, non-verbal IQ score, and positive and negative false alarms. For additional confounders, model 2 was additionally adjusted for ethnicity, English as a first language, dyslexia, autism spectrum disorders, parental education, maternal depression, and paternal depression. For pubertal stage, model 2 was additionally adjusted for pubertal stage. Negative binomial mixed models were used in the subsample with data on additional confounders, and Poisson mixed models were used in the subsample with data on pubertal stage.

Supplementary Table 4. *Negative binomial and Poisson mixed models testing interactions between age group, gender, condition (self-/other-referential) and valence (positive/negative) on total hits, in subsamples with additional confounders or pubertal stage reported.*

	Subsample with data on additional confounders N=275		**Subsample with data on pubertal stage** N=117	
	Column 1: Adjusted	**Column 2: Additionally adjusted**	**Column 1: Adjusted**	**Column 2: Additionally adjusted**
	Interaction term p value		Interaction term p value	
Gender x condition	0.56	0.56	0.10	0.10
Gender x valence	0.90	0.90	0.95	0.95
Gender x condition x valence	0.67	0.67	0.83	0.83
Age group x gender	0.23	0.29	0.15	0.12
Age group x gender x condition	0.15	0.15	0.09	0.09
Age group x gender x valence	0.16	0.16	0.41	0.41
Age group x gender x condition x valence	0.91	0.91	0.79	0.79

Note. Each interaction term was tested in a separate model. All models adjusted for condition, valence, gender, age group, continuous age within each group, school, testing group size, non-verbal IQ score, and positive and negative false alarms. For additional confounders, column 2 was additionally adjusted for ethnicity, English as a first language, dyslexia, autism spectrum disorders, parental education, maternal depression, and paternal depression. For pubertal stage, column 2 was additionally adjusted for pubertal stage. Negative binomial mixed models were used in the subsample with data on additional confounders, and Poisson mixed models were used in the subsample with data on pubertal stage.

Supplementary Table 5. *Linear regression models testing change in depressive symptoms (SMFQ score) for each additional self-referential positive, self-referential negative, other-referential positive, and other-referential negative hit, in subsamples with additional confounders or pubertal stage reported.*

	Subsample with data on additional confounders N=275				**Subsample with data on pubertal stage** N=117			
	Model 1: Adjusted		**Model 2: Additionally adjusted**		**Model 1: Adjusted**		**Model 2: Additionally adjusted**	
	Coef (95% CI)	p value	Coef (95% CI)	p value	Coef (95% CI)	p value	Coef (95% CI)	p value
Self-referential hits								
Positive	-0.35 (-0.75 to 0.04)	0.08	-0.29 (-0.69 to 0.11)	0.16	-0.27 (-0.96 to 0.42)	0.44	-0.27 (-0.96 to 0.42)	0.44
Negative	0.66 (0.23 to 1.09)	0.003	0.65 (0.20 to 1.10)	0.005	0.51 (-0.23 to 1.24)	0.18	0.52 (-0.22 to 1.26)	0.17
Other-referential hits								
Positive	-0.01 (-0.45 to 0.42)	0.95	-0.01 (-0.45 to 0.43)	0.96	-0.02 (-0.80 to 0.75)	0.95	-0.06 (-0.84 to 0.72)	0.88
Negative	0.06 (-0.37 to 0.49)	0.79	0.05 (-0.39 to 0.49)	0.84	0.48 (-0.26 to 1.23)	0.20	0.55 (-0.21 to 1.31)	0.16

Note. Models included all four types of hits as independent variables and were adjusted for age group, gender, continuous age within each group, school, testing group size, non-verbal IQ score, and positive and negative false alarms. For additional confounders, Model 2 was additionally adjusted for ethnicity, English as a first language, dyslexia, autism spectrum disorders, parental education, maternal depression, and paternal depression. For pubertal stage, Model 2 was additionally adjusted for pubertal stage.