

Supplementary information

Supplementary tables and figures

Task	Feature	Description
Verbal fluency task: Controlled Oral Word Association Test (COWAT)	Vocabulary score	A score designed to measure the average word frequency of words said by the participant. Word frequencies were computed in sample. This score is the $-\log_{10}(\text{mean word frequency})$. High scores indicate the participant is consistently using words most other people are not, and may indicate a high vocabulary.
	Unique word count	Number of unique words in a response. (Estimate for verbal fluency)
	Mean time to say word	Average time taken to say a word, a proxy of word length and verbal fluency
	Hapax word count	A hapax is a word that was said only once in our sample. This score represents the total number of hapaxes said by the participant.
Expressive language task: Picture narration	Unique word count	Number of unique words in narratives by a participant
	Mean time to say word	Average time taken to say a word, a proxy of word length and verbal fluency
	Hapax word count	A hapax is a word that was said only once in our sample. This score represents the <u>total</u> number of hapaxes said by the participant.
Rhythm/timekeeping task: Sync voice	Consistency (R-squared)	Participant-level R-squared values from linear models to predict accuracy
	Mean error	Participant-level mean error to predict accuracy
Phonological memory task: Sentence repetition	Minimum bigram accuracy	Minimum accuracy of bigrams (consecutive word pairs) in all trials
	Mean bigram accuracy	Average of bigram accuracy.
Nonverbal reasoning task: Matrix reasoning	Proportion correct	Proportion of correct trials
Spatial receptive language task: Following directions	Proportion correct	Proportion of correct trials
Reading task: Rapid automatized naming (RAN)	Time to complete	Average time it took participants to complete the RAN trials (in seconds)

Table 2. Description of Lingo features.

prompt number	prompt
1	didnt the girl paint the picture
2	the hamster was not left in the cage
3	the student did not see the teacher who taught first grade last year
4	if the driver had delivered some donuts they would have been eaten by now
5	the woman who helps serve the hot lunches at school is the new nurse
6	the girl who didnt participate in practice wasnt allowed to start for the team until a month later
7	the art teacher prepared printed organized and delivered the new sketchbooks
8	the janitor from down the hall promised to feed the python during our winter vacation
9	it is the teachers practice to always assign homework two days before it is due and no more than three days following the previous assignment
10	the cafeteria doors around which the seventh graders sometimes stand until the lunch bell rings are always closed but never locked
11	by secretly helping herself to his extra pencils she was ensuring neither would have one available when the big test he had been worrying about was finally given
12	having been tardy more days than not the brothers both knew it was only a matter of time before each would have to be responsible for himself

Figure S1. Prompt sentences used in the sentence repetition task (with punctuation and capitalization removed).

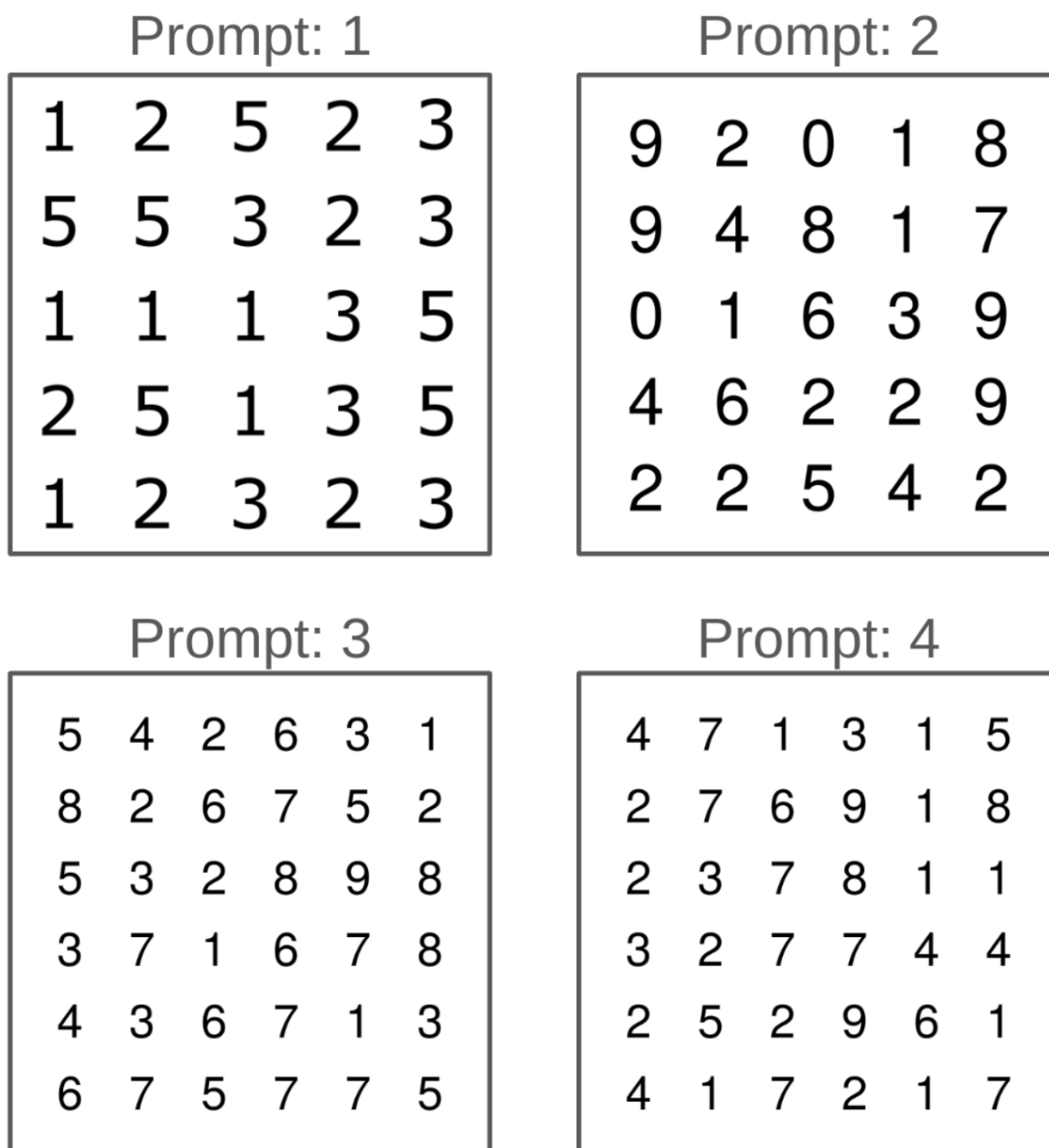


Figure S2. Prompt number grids used in the RAN reading task.

Figure S3. Please contact the authors for a copy of the prompt images. Prompt images used in the picture narration task will be included here in the final peer reviewed version of the manuscript.

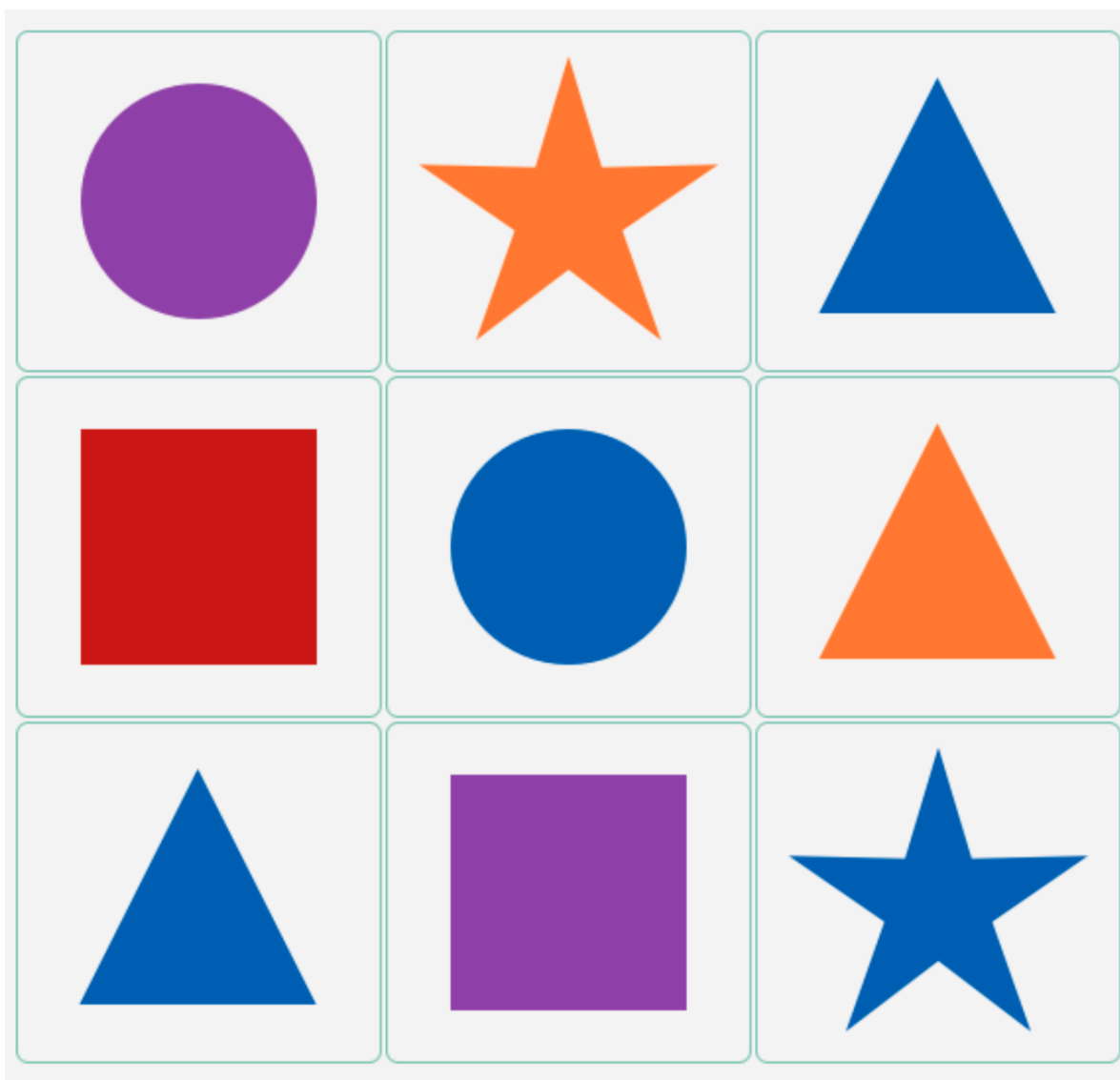


Figure S4. Prompt 1 image for the following directions. Participants were asked to click on the shape that matched the verbal description they had heard.

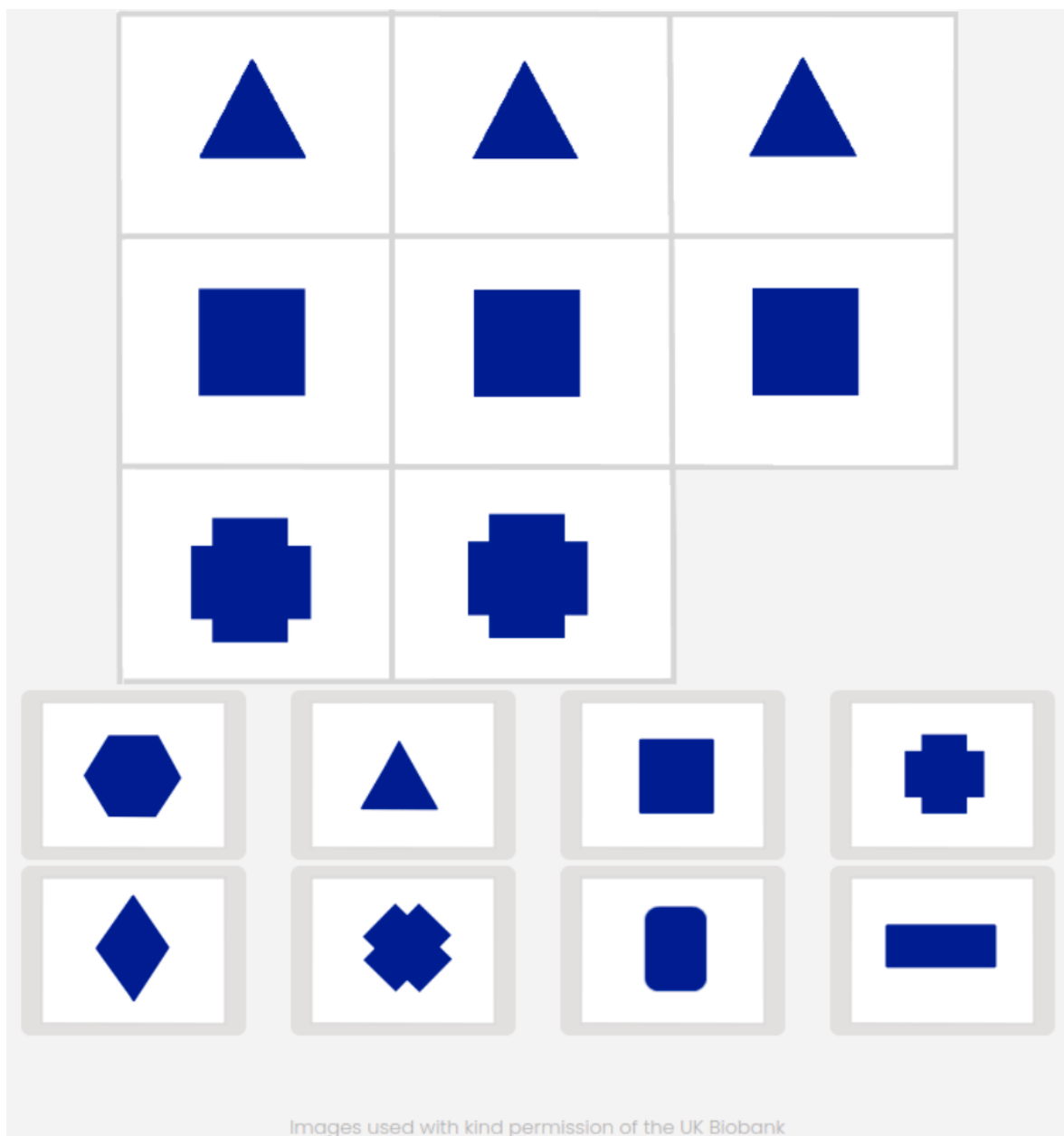


Figure S5. Prompt 1 image for the matrix reasoning task, used with permission from same task in the UK Biobank. Participants were asked to click on the pattern that best fit the matrix shown above.

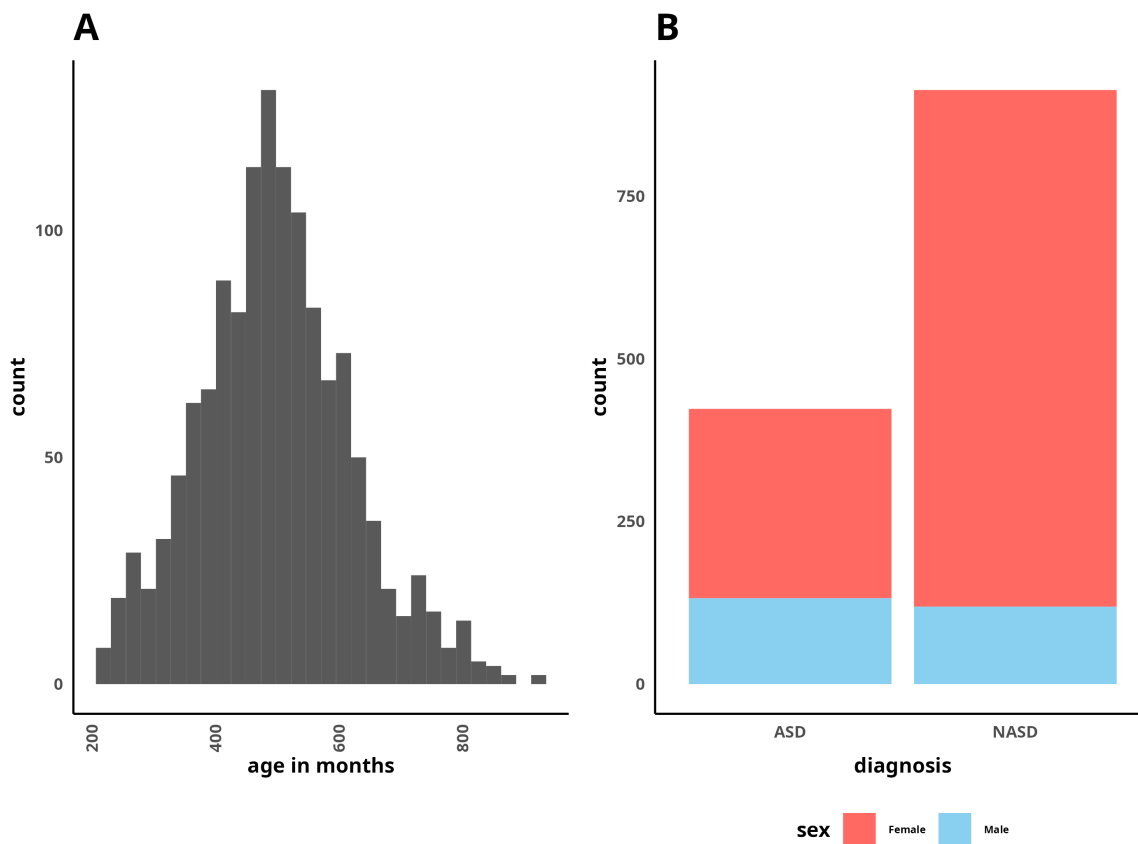


Figure S6. Demographic distributions of the SPARK Lingo cohort.