

Mothers' speech predicts children's differences in cognition, literacy, and educational achievement across the school years

Anna Brown^{*1}; Avshalom Caspi^{2,3,4,5,6,7}; Helen L. Fisher^{2,3}; Terrie E. Moffitt^{2,3,4,5,6,7}; Jasmin Wertz⁸; Sophie von Stumm¹

1. Department of Education, University of York, York, UK
2. King's College London, Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, London, UK
3. ESRC Centre for Society and Mental Health, King's College London, London, UK
4. Department of Psychology and Neuroscience, Duke University, Durham, NC, USA
5. Department of Psychiatry and Behavioral Sciences, Duke University, Durham, NC, USA
6. Duke University Population Research Institute, Duke University, Durham, NC, USA
7. PROMENTA, Department of Psychology, University of Oslo, Oslo, Norway
8. Department of Psychology, University of Edinburgh, Edinburgh, UK

Corresponding Author: Anna Brown; Department of Education, York, Heslington, YO10 5DD, UK;

Email: anna.brown2@york.ac.uk

Author Contribution: AB, AC, HLF, TEM, JW, and SvS designed and carried out the research; AB and SvS analyzed the data and wrote the paper. All authors critically reviewed and edited the content of the draft manuscript and provided final approval before submission.

Competing Interest Statement: The authors declare that they have no competing interests.

Data and Code Availability: All data are available upon request to the E-Risk study steering committee (<https://eriskstudy.com/data-access/>). The analysis code for this study is available from the corresponding author upon request.

Abstract

Mothers' speech is likely a key driver of family background inequality in education. However, long-term associations between mothers' speech and children's educational outcomes beyond early years have not been previously tested. Here, we quantified differences in mothers' vocabulary sophistication, lexical diversity, and grammatical complexity from 10-minute-long excerpts from audio-recorded interviews when children were 5 years old and starting primary school. Families represented the full range of Britain's socioeconomic conditions (analytic sample $N = 894$ mother-child trios). Mothers' vocabulary sophistication significantly predicted children's cognition, literacy, and educational achievement between the ages of 5 and 12 years, accounting for 2% to 5% of the variance. These estimates were reduced to 1% and 2% or became altogether non-significant after adjusting for mothers' educational attainment and household income. Our findings suggest that mothers' speech plays a small independent role in the long-term transmission of family background inequality in education.

Introduction

Doing well in school is the first step to success in meritocratic societies that allocate resources and power according to demonstrated achievement. A child's school performance may forecast the number, level, and prestige of the educational qualifications they will attain throughout their lifetime^{1–3}. These educational qualifications will, in turn, inform the child's later access to economic, social, and cultural resources⁴, their position in society^{5,6}, well-being⁷, health⁸, and life expectancy⁹. Yet, children start formal education with unequal opportunities for doing well in school. Children from families with fewer socioeconomic resources tend to lag in school readiness and achieve lower grades in primary and secondary education than children from well-resourced families^{10,11}.

Prior studies identified the speech children hear from their mothers as a key pathway through which family background inequality in education becomes transmitted^{12–15}. Mothers from under-resourced families speak, on average, less frequently to their children¹⁶, using less diverse vocabulary and simpler syntax than mothers from well-resourced families^{17–19}. Mothers' child-directed lexicon and grammar during the early years predict children's linguistic development (e.g., word segmentation and word mapping)²⁰ and their school readiness²¹. Yet, little is known about the role of mothers' adult-directed speech, which supersedes child-directed speech by the time children reach the age of 5 years and start formal education^{21,22}. To address this gap, we report the long-term prediction of children's cognition, literacy, and educational achievement at age 5 through to 12 years from mothers' vocabulary sophistication, lexical diversity, and grammatical complexity. We use the term "prediction" in this article to refer to the temporal order in which the assessment of one variable preceded the assessment of another.

Mothers' speech may contribute to achievement gaps in children's education via two principal mechanisms^{23,24} that likely intersect. For one, mothers' speech is thought to benefit children's growth in developmental domains that thrive through exposure to language, such as school performance²⁵, the ability to reason and argue logically²⁶, and literacy skills²⁷. Alternatively, mothers' speech may account for children's degree of familiarity with the linguistic codes that prevail in formal education settings^{28,29}. Children raised without many opportunities to learn these linguistic codes or styles may – unwittingly – communicate in language patterns that lead to stereotyping, ridicule, discrimination, and exclusion^{30,31}.

Our study elucidated the role mothers' speech likely plays in transmitting family background inequality in education in four ways. First, we modeled the long-term predictions of mothers' speech for their children's education-related outcomes across the school years. Second, we tested whether mothers' socioeconomic status (SES) was associated with the characteristics of their speech. Third, we explored the extent to which predictions from mothers' speech for children's education-related outcomes were independent of mothers' SES, and then fourth, if mothers' speech mediated the association between mothers' SES and children's education-related outcomes.

We capitalized on rich, naturalistic speech data collected from a socioeconomically diverse sample of $N = 894$ mothers when their twin children were 5 years old and starting primary school³². Data came from E-Risk, a UK population-representative cohort study of families who had twins born from 1994 to 1996 in England and Wales³³. The mother-child trios in E-Risk were identified via the Office for National Statistics' multi-birth register³³ and represent the full range of Britain's socioeconomic conditions. E-Risk families' residential addresses match the UK Index of Multiple Deprivation distribution: 25% of the families lived in Britain's wealthiest and 25% in the poorest

areas ³⁴. Because E-Risk has suffered minimal attrition since its inception (~5%), in contrast to many other longitudinal studies ³⁵, the sample's representativeness remained intact over time.

Mothers freely described their parenting experiences and the twins' characteristics in interviews conducted in person by trained health researchers ³² (i.e., adult interviewers; average interview length ~22 minutes). From excerpts of these interviews' transcripts, we quantified mothers' lexicon in terms of vocabulary sophistication (% of rare words used in speech ³⁶) and lexical diversity (range of word types used in speech ³⁷). We also quantified mothers' grammatical complexity from the mean length of utterance, verb usage (number of verbs per utterance ³⁸) and syntactic complexity (proportion of grammatical relations reflecting complex syntactic structures ³⁹). For descriptives and correlations of lexical and grammar markers, see Supplementary Table 1 and Supplementary Fig. 1. Children's verbal and non-verbal cognitive abilities at age 5 years and literacy skills (i.e., reading ability) at age 7 and 10 years were assessed with standardized psychometric tests. Children's performance in the core school subjects of English and Mathematics at age 7, 10, and 12 years were rated by teachers relative to that of other pupils at the same age. Using latent growth curve models, we differentiated the variance in children's achievement that was stable from primary through secondary school from the variance in achievement that reflected children's systematic performance gains and losses as they progressed through the school years ⁴⁰. This approach allowed testing whether mothers' speech produced dynamic changes in educational achievement during the school years or if mothers' speech was associated primarily with children's differences in educational achievement that remained unchanged throughout primary and secondary school.

Results

We fitted hierarchical linear regression models and latent growth curve models using the R package Lavaan ⁴¹ to test if mothers' speech was associated with their SES and predicted their children's cognition, literacy, and educational achievement. To account for the non-independence of observations (i.e., twins), standard errors were clustered at the family level; mothers' age was included as a covariate in all models. E-Risk incurred very low attrition over time (~5%) with data missing randomly; we used Full Information Maximum Likelihood (FIML) in all models ⁴².

Predicting children's cognition and literacy from mothers' speech

We tested the extent to which mothers' speech predicted children's cognition and literacy concurrently and prospectively at the ages of 5, 7, and 10 years, assessed by standardized psychometric tests ^{43,44}, which in turn forecast their educational achievement ²². We fitted independent models for each measure at each age with mothers' speech characteristics as predictors.

Mothers' vocabulary sophistication was significantly and positively associated with children's cognition and literacy (Supplementary Tables 2-4), accounting for 2-5% of the variance across ages and measures (Fig. 1a). Neither mothers' lexical diversity nor grammar significantly predicted children's cognition and literacy (Fig. 1a).

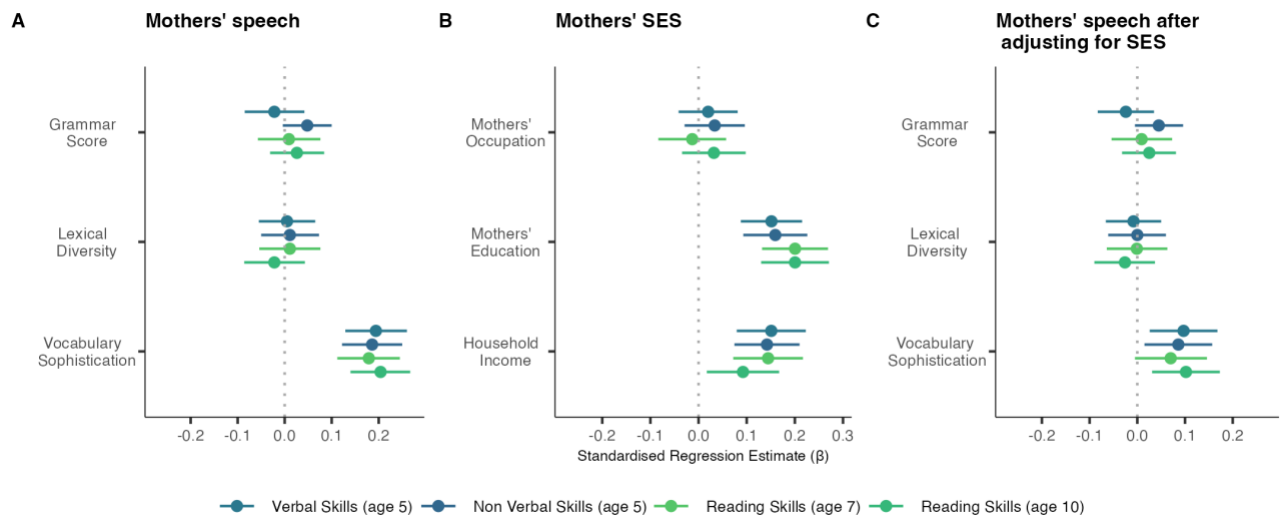


Fig 1. Predicting children's cognition and literacy from mothers' speech and SES.

Note. Plot A (left) shows the associations between mothers' speech and children's cognition and literacy; Plot B (middle) shows associations between mothers' socioeconomic status (SES) and children's cognition and literacy; Plot C (right) shows the association between mothers' speech and children's cognition and literacy after adjusting for SES. Dots represent standardized (beta) regression estimates, and lines represent 95% confidence intervals. All models were adjusted for mothers' age.

Predicting children's educational achievement from mothers' speech

We next tested the extent to which mothers' speech predicted children's school performance in English and Mathematics, as rated by teachers when the children were aged 7, 10, and 12 years. We used latent growth curve models to discern associations between mothers' speech and children's variance in educational achievement that was stable throughout the school years (intercept) versus variance in educational achievement that reflected systematic gains and losses in school performance over time (slope; see Supplementary Note 1). This approach enabled testing whether mothers' lexicon and grammar predicted differences in children's educational achievement that were stable across the school years or predicted dynamic changes in educational achievement over time.

We specified mothers' speech characteristics (i.e., vocabulary sophistication, lexical diversity, and grammatical complexity) as predictors of the two latent growth factors (i.e., intercept and slope). Mothers' vocabulary sophistication was significantly associated with children's stable variance in educational achievement (intercept), accounting for 2% of the variance, but not with gains and losses in educational achievement over time (slope; Supplementary Table 5). Thus, mothers' vocabulary sophistication predicted children's achievement differences that persisted throughout school with a small effect size but did not forecast children's improvements or worsening in school performance over time relative to where they started. Mothers' lexical diversity and grammatical complexity were associated neither with the intercept nor the slope of children's educational achievement (Supplementary Table 5). Akin to our findings on children's cognition and literacy, these results suggest that mothers' vocabulary sophistication predicts children's differences in educational achievement. Yet, mothers' vocabulary sophistication was not associated with changes in achievement trajectories that occur over time.

Mothers' SES and characteristics of their speech

To explore the role mothers' speech plays in the transmission of family background inequality, we first tested whether mothers' SES, including their educational attainment, occupation, and household income, was associated with their speech ⁴⁵. We fitted independent regression models for each characteristic of mothers' speech (i.e., vocabulary sophistication, lexical diversity, and grammatical complexity) to be predicted by mothers' SES indicators.

Mothers' SES indicators accounted for 28% of the variance in vocabulary sophistication and for 6% of the variance in lexical diversity; they were not significantly associated with grammatical complexity (Fig. 2, Supplementary Table 6). These results suggest that only measures of mothers' lexicon (i.e., vocabulary sophistication and lexical diversity) vary as a function of their SES.

Mothers' educational attainment was the strongest independent correlate of vocabulary sophistication, sharing 9% of the variance, followed by household income and occupation, which each accounted for 5% of the variance. Household income also accounted for 4% of the variance in lexical diversity, but no other significant SES associations with mothers' speech were observed (Supplementary Table 6). Sensitivity analyses (Supplementary Note 2) showed that the effect sizes were robust after controlling for mothers' scores on the reading subtest of the wide-range achievement test (WRAT-3 ⁴⁶). Overall, the findings suggest that mothers' SES is more strongly

and consistently associated with their vocabulary sophistication than their lexical diversity and grammar.

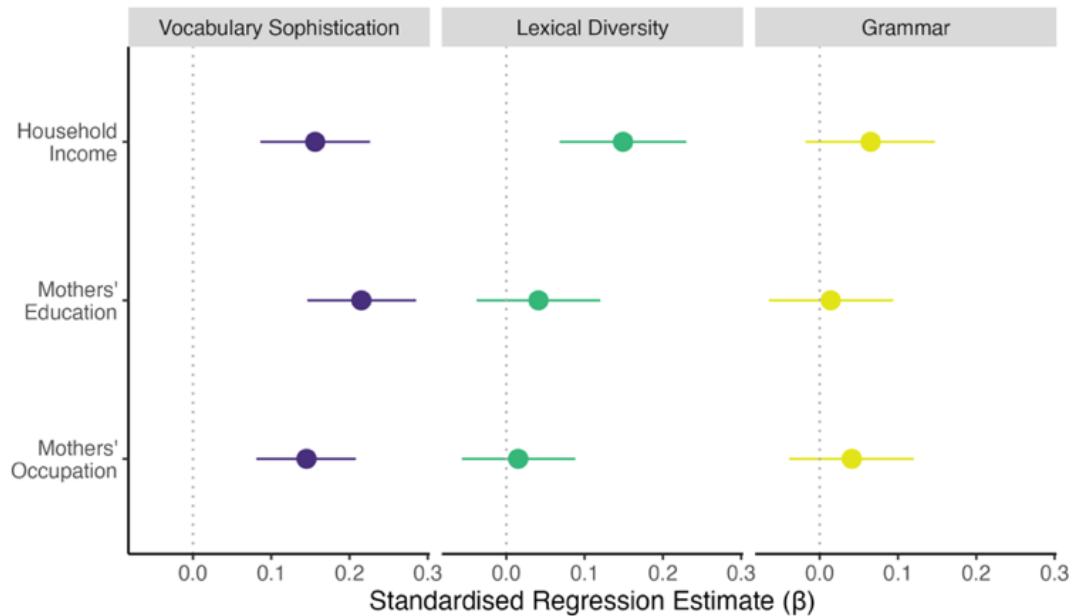


Fig. 2. Mothers' education, occupation, and household income as predictors of their speech characteristics

Note: N = 894. Dots represent standardized regression (beta) estimates from models where mothers' education, occupation, and income were included as simultaneous predictors after adjusting for mothers' age in years. The lines indicate the 95% confidence intervals.

Predicting children's cognition and literacy from mothers' SES

Next, we modeled the extent to which mothers' SES predicted children's cognition and literacy (Fig. 1b; Supplementary Tables 2-4). Across ages and measures, mothers' SES indicators accounted for 8-10% of the variance in children's cognition and literacy (Fig. 1b). Mothers' educational attainment and household income accounted independently for 3-4% of the variance in children's cognition and literacy across ages and measures. Mothers' occupation was not significantly associated with children's cognition and literacy at any age.

Predicting children's educational achievement from mothers' SES

Mothers' educational attainment and household income, respectively, accounted for 9% and 14% of children's stable variance in educational achievement (Supplementary Table 5) but were not associated with changes in achievement (slope) across the school years. Mothers' occupation was not significantly associated with children's educational achievement. Our findings suggest that mothers' educational attainment and household income can predict children's differences in

performance that are already evident at the start of schooling but not relative changes in achievement.

Discerning predictions from mothers' speech versus SES for children's cognition and literacy

Next, we tested the extent to which our previously observed associations between mothers' vocabulary sophistication and children's cognition and literacy could be attributed to SES (Fig. 1; Supplementary Tables 2-4). Associations between mothers' vocabulary sophistication and children's cognition and literacy remained significant but halved in effect size (i.e., 1-2%) after adding mothers' SES to the models. These findings suggest that associations of mothers' SES with children's cognition and literacy are likely more pervasive than those with mothers' vocabulary sophistication. Our sensitivity analyses (see Supplementary Note 2) confirmed that the observed associations were robust and independent of covariate effects. We observed no differences in the associations between mothers' lexical diversity and grammar with children's cognition and literacy, which were non-significant before and after adding mothers' SES to the models.

To explore if mothers' vocabulary sophistication mediated associations between SES and children's cognition and literacy (Fig. 3A), we specified mothers' vocabulary sophistication as a mediator and mothers' education and household income as predictors (Fig. 3A). The latter were significantly associated with children's cognition and literacy in our earlier analyses, which is a pre-requisite for fitting mediation models ⁴⁷ (see Supplementary Note 3 for full details).

Mothers' vocabulary sophistication significantly mediated the association between mothers' educational attainment and children's verbal ($\beta_m^* = 0.04$, 95% CI= 0.01-0.07) and non-verbal skills at age 5 ($\beta_m = 0.03$, 95% CI= 0.01- 0.05) as well as their literacy skills at age 10 ($\beta_m = 0.03$, 95% CI= 0.01-0.06). At age 7, the mediation for literacy skills was not significant ($\beta_m = 0.02$, 95% CI= -0.01-0.05; Fig. 3B and Supplementary Table 7). Mothers' vocabulary sophistication significantly mediated the association between household income and children's verbal ($\beta_m = 0.03$, 95% CI= 0.01-0.05) and non-verbal skills at age 5 ($\beta_m = 0.02$, 95% CI= 0.01- 0.04), and their literacy skills at age 10 ($\beta_m = 0.02$, 95% CI= 0.01-0.05), but not at age 7 ($\beta_m = 0.02$, 95% CI= -0.01-0.05).

Effect sizes for the significant mediations were small, ranging from 13% to 24% of the associations between SES and children's cognition and literacy, with models accounting for 11% to 8% of the variance (i.e., absolute mediation effect sizes were 3% to 1%; full model results in Supplementary Table 7). These results suggest that mothers' vocabulary sophistication partly explains the associations between mothers' educational attainment and household income with children's cognition and literacy, albeit with modest effect sizes and inconsistently across measures.

* β_m refers to product of the paths (predictor \rightarrow mediator) * (mediator \rightarrow outcome) that specifies the indirect effect estimate (i.e., mediation effect; blue arrows in Fig. 3A).

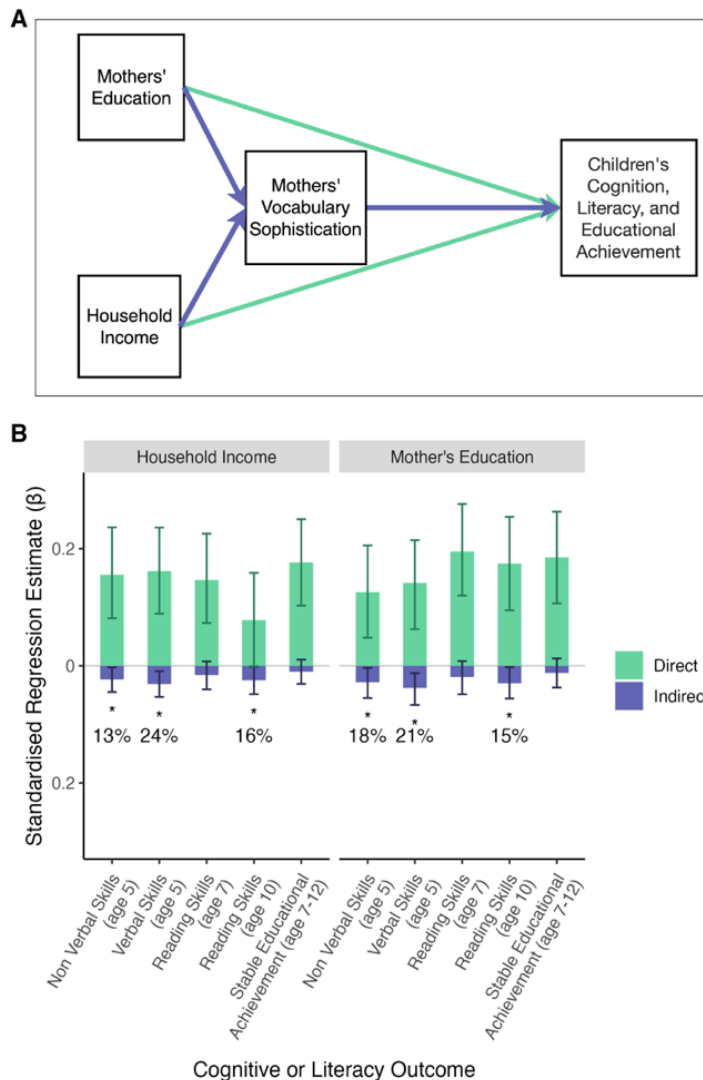


Fig. 3. Direct and mediated effects of mothers' vocabulary sophistication on children's cognition, literacy, and educational achievement.

Note. $N=894$, Top panel (A) illustrates the principal mediation model that was fitted. Green arrows represent direct effects; blue arrows represent mediated effects. Bottom panel (B) shows the direct and mediated estimates (betas transformed into % of variance explained by indirect effect) with 95% bootstrapped confidence intervals from 1000 resamples for mothers' vocabulary sophistication (mediator), separately for mothers' education and household income (predictors) cognition, literacy, and educational achievement (outcomes).

Discerning predictions from mothers' speech versus SES for children's educational achievement

We tested the extent to which associations between mothers' vocabulary sophistication and children's stable variance in educational achievement could be attributed to SES (Supplementary Table 6). Associations between mothers' vocabulary sophistication and children's educational achievement became non-significant after adding mothers' SES to the models, suggesting again

that associations between mothers' SES and children's educational achievement are likely more pervasive than those with mothers' vocabulary sophistication.

Because mothers' vocabulary sophistication was independently a significant predictor of children's stable variance in educational achievement, the pre-requisite for testing mediation was met. We specified mothers' vocabulary sophistication to mediate the associations between their educational attainment or household income and their children's educational achievement. Mothers' vocabulary sophistication mediated neither association (educational attainment : $\beta_m = 0.01$, 95% CI = -0.01-0.04; household income: $\beta_m = 0.01$, 95% CI = -0.01-0.03; Fig. 3B; Supplementary Table 7). Above, we found that mothers' vocabulary sophistication mediated the relationship between mothers' SES and children's cognition and literacy. Yet, for educational achievement, the mediations were not significant. This discrepancy in results suggests that mothers' vocabulary sophistication may play a different role in children's development of cognition and literacy as compared to their educational achievement.

Discussion

The language environments children experience during their early years likely drive the transmission of family background inequality in education^{13,48}. Previous research conceptualized these early life language environments primarily in terms of mothers' child-directed speech^{15,19}, which facilitates children's initial acquisition of language²⁰. Many empirical studies reported significant associations between mothers' child-directed speech and children's vocabulary^{17,19,36} and grammar¹⁸ during the early years. Yet, our study is the first to test whether mothers' adult-directed speech predicts children's long-term cognition, literacy, and educational achievement across primary and secondary school years.

Our study's findings may explain why interventions that target mothers' speech to disrupt family background inequality in education often show only short-term benefits^{49,50}. In 894 families drawn from a UK-representative longitudinal cohort study, we observed that mothers' vocabulary sophistication accounted for 2% of the variance in children's educational achievement across school years. This small effect became non-significant after taking into consideration mothers' educational attainment and household income, suggesting that mothers' vocabulary sophistication does not share pervasive associations with children's educational achievement. Educational attainment and household income reflect mothers' relative access to and availability of socioeconomic resources^{51,52} and ability to invest in child-specific learning activities⁵³. Respectively, they explained 9% and 14% of children's differences in educational achievement, a fivefold greater effect than that observed for mothers' vocabulary sophistication.

We contend that mothers' child-directed speech predicts children's language abilities during the early years^{19,54,55} and that these early language abilities aid children's education-related outcomes^{21,22}. However, our findings here and elsewhere⁵⁶ challenge the hypothesis that mothers' speech exerts significant influence on children's educational achievement after they have started formal education. Past research on the transmission of family background inequality in education may have focused too narrowly⁵⁷ on the role of mothers' speech at the cost of studying alternative linguistic¹² and non-linguistic pathways^{58–60} that likely carry equal, if not greater, explanatory power. Intervention efforts based on this narrow research may have attributed – inadvertently – too much responsibility to mothers' speech for children's educational successes and failures⁶¹. Such misattributions can reinforce narratives that mothers, rather than the social and economic conditions in which they raise their children, are at the root of children's educational achievement differences^{57,62,63}.

Our findings also challenge the idea that mothers' speech differences are shaped by the 'social gradient', a widely observed phenomenon in which people with fewer socioeconomic resources are at increased risk of less favorable outcomes (e.g., worse speech) relative to people with greater resources⁶⁴. We found small associations between mothers' SES and their vocabulary sophistication and lexical diversity, and no significant association with their grammatical complexity. These findings align with analyses of other large-scale cohorts^{65,66} and meta-analytic studies^{67,68} but contradict some earlier reports of dramatic SES-related differences in mothers' speech in smaller samples^{16,18,69}. Characteristics of mothers' speech other than the ones studied here may converge on the social gradient, such as dialect⁷⁰, word pronunciation⁷¹, and narrative construction⁷². Yet, we found little evidence in the current analyses that vocabulary sophistication, lexical diversity, and grammatical complexity, which are typical markers of adult-directed speech, covaried systematically with education, occupation, and household income that are typical markers of SES.

Our overall finding was that mothers' speech likely plays a modest role in children's cognition, literacy, and educational achievement beyond the early years. In addition, two notable observations emerged. First, among the three characteristics of mothers' speech studied here, vocabulary sophistication (i.e., % of rare words³⁶) was the most consistent predictor of children's

education-related outcomes. It was also the speech characteristic most strongly associated with mothers' SES. The potential influence of mothers' vocabulary sophistication on children's vocabulary development has been previously noted ^{36,73}. However, the exact pathways and mechanisms through which mothers' vocabulary sophistication may inform children's developmental differences in education-related outcomes are, to date, unknown ^{74,75}. That said, vocabulary sophistication is believed to benefit children's learning from academic texts, which become increasingly abstract, complex, and dense across school years ^{76,77}. Thus, our findings lend some support to calls for classroom environments that encourage engaging with sophisticated vocabulary that is aligned with the complexity of the academic texts that pupils study ⁷⁸.

Second, we found that mothers' vocabulary sophistication predicted children's cognition and literacy more strongly than children's educational achievement in two ways. For one, mothers' vocabulary sophistication continued to predict children's cognition and literacy after considering mothers' educational attainment and household income, while its prediction of educational achievement became nonsignificant. For the other, mothers' vocabulary sophistication partly mediated associations between mothers' educational attainment and household income with children's cognition and literacy but not with children's educational achievement. Cognition and literacy are narrow skills that inform educational achievement ^{79,80}. Educational achievement is a broad skill encompassing many other factors, such as school attendance, study habits, and classroom behavior ^{81,82}. Mothers' vocabulary sophistication may have direct benefits for the development of cognition and literacy ^{79,83,84} that we could observe here. By comparison, its influences on educational achievement may be more indirect ⁸⁵ and, thus, were masked in the current analyses.

Our study has many strengths, including rich observational data across a 7-year follow-up period from a large sample representative of the full range of Britain's socioeconomic conditions ⁸⁶. Yet, it is not without limitations. First, because no measures of mothers' child-directed speech were available, comparing the lexical and grammatical complexity in mothers' child- versus adult-directed speech ^{17,19,87} and respective associations with children's education-related outcomes was impossible. Child-directed speech is likely pivotal for children's language acquisition during the early years ^{14,16}, but its long-term influences on children's educational trajectories have not been studied, a gap that future research should address. Second, our study focused only on mothers' speech but no data were available on the speech of fathers, teachers, and peers, which may also influence children's language development and education-related outcomes ^{88–90}. Thus, our analyses allow inferring conclusions about the role of mothers' speech in the transmission of family background inequality in education but not about children's broader speech environments. Third, while our study's data are comparatively rich, they were collected infrequently (i.e., at varying time windows between 2 to 5 years) up to the beginning of secondary school, assessing some learning skills (i.e., cognition and literacy) but not others (e.g., self-regulation, numeracy). We therefore caution that, while our findings are clear and consistent, future research needs to replicate and extend them.

Our findings suggest that mothers' education and household income can exert greater long-term influence on their children's education-related outcomes than the characteristics of mothers' speech. Intervention efforts that seek to disrupt the perpetuation of family background inequality in education should consider a broad range of possible transmission pathways, rather than attributing the responsibility for children's educational outcomes disproportionately to how mothers speak. Pedagogical approaches at the classroom level ⁹¹ that enable all children to maximize the learning opportunities that school affords them, regardless of their cognition and literacy skills at school entry or their families' socioeconomic resources, may be more effective for reducing family background inequality in education ^{63,92} than modifying children's home language experiences. A first step to effectively reducing family background inequality in education is

recognizing that achievement gaps emerge and manifest in the context of wider, systemic discriminations.

Materials and Methods

Sample

Data came from the Environmental Risk (E-Risk) Longitudinal Twin Study, which tracks the development of 2,232 British twin children drawn from a larger birth register of twins born in England and Wales in 1994 and 1995⁹³. Full details about the sample are reported elsewhere³³ and in Supplementary Note 4. To summarize, the E-Risk sample was constructed in 1999 to 2000 when 1,116 families (93% of those eligible and of whom 90.4% were Caucasian) with same-sex 5-year-old twins participated in home-visit assessments. This sample comprised 56% monozygotic and 44% dizygotic twin pairs; sex was evenly distributed within zygosity (49% male). Families were recruited to represent the UK population of families with newborns in the 1990s, on the basis of residential location throughout England and Wales and mothers' age (i.e., older mothers having twins via assisted reproduction were under-selected and teenage mothers with twins were over-selected). This resulted in a sample of families which were representative of the distribution of neighborhood socio-economic status across the UK population⁹⁴.

Follow-up home visits were conducted when children were aged 7, 10, 12, and 18 years (participation rates were 98%, 96%, 96%, and 93%, respectively). Here we use data collected up to age 12 when 1073 families participated. Home visits at ages 5, 7, 10, and 12 years included assessments with participants and their mother (or primary caregiver); a different interviewer assessed each participant in a twin pair. With parents' permission, questionnaires were posted to the children's teachers, who returned questionnaires for 94% of children at age 5, 91% of the 2,232 E-Risk children (93% of those followed up) at age 7, 86.3% of the 2,232 E-Risk children (90.1% of those followed up) at age 10, and 80% of the 2,232 E-Risk children at age 12 (83% of those followed up). The Joint South London and Maudsley and Institute of Psychiatry Research Ethics Committee approved each study phase. Parents gave informed consent and twins gave assent between 5 and 12 years and then informed consent at age 18.

Mothers' speech corpus

When their twin children were 5 years old, mothers were asked to speak for 5 to 10 minutes about each of the children separately during home visit assessments³². Mothers were encouraged to talk freely with few interruptions. However, if a mother found this difficult, the interviewer could aid her with a series of semi-structured probes, such as 'In what ways would you like [child] to be different?'. Interviews about each twin were separated in time by approximately 90 minutes. All interviews were audiotaped with the mothers' consent. Interview data was missing for 6% of the sample because some mothers did not wish to be audiotaped, because of technical problems with the tape, or because of loss to follow-up. The audio-taped interviews were transcribed, and 10-minute samples from each interview (to ensure comparability given the varying interview lengths) were coded for automatic analyses using the Computer Language Analysis (CLAN) program⁹⁵. A further 1% of interviews were excluded at the transcription stage due to poor audio. 2% of families were then excluded because a language other than English was spoken at home. Our final analysis sample included all mothers and their twins for whom sufficient speech data was available (i.e., 1,788 children from 894 families).

Mothers' speech characteristics

From the transcripts of mothers' speech, we extracted quantitative markers of lexicon and grammar using the *FREQ*, *VOCD*, and *MLU* programs within CLAN⁹⁵. For lexical markers, we extracted mothers' vocabulary sophistication and lexical diversity. For grammar markers, we extracted mothers' mean length of utterance (morpheme count), grammatical complexity index, and verb usage within utterances. When extracting grammar markers, partially unintelligible utterances were excluded.

Vocabulary sophistication

Vocabulary sophistication was the percentage of rare words mothers used in their respective speech transcripts. From the transcripts, we removed all words on the Dale-Chall word list (79), their inflected forms, all non-dictionary words, and any derivations of family names. The Dale-Chall word list ⁹⁶ includes the 3,000 most common words known by fourth graders. Any remaining words in the transcripts were considered rare and sophisticated word types ³⁶. The percentage of mothers' rare word types relative to their total number of word types in the transcript was then calculated.

Lexical diversity

Lexical diversity was measured by computing D scores, which index the probability of introducing a new word as the speech sample progresses ⁹⁵. D-scores are highly robust against sample size effects and thus constitute reliable markers of lexical diversity over other methods such as Type Token Ratio ^{97,98}.

Grammatical complexity

Grammatical complexity was inferred from the grammatical complexity index (GCI) ³⁹, verb index, and mean length of utterance ³⁸.

GCI was computed by calculating mothers' total number of grammatical relations that marked syntactic embeddings divided by their total number of grammatical relations ^{39,99} — the higher the index, the more complex grammatical and syntactic structures used in speech. The verb index was the mothers' total number of verbs divided by the total number of utterances. The higher the average number of verbs per utterance, the more complex the syntax. MLU was calculated by dividing mothers' total number of morphemes by their total number of utterances produced. The higher the MLU, the more complex and longer a person's utterances. Any words classed as communicators, such as 'oh' or 'ah', were excluded from this analysis.

Preliminary analyses were carried out to validate the reliability of the characteristics of mothers' speech. Our lexical markers (vocabulary sophistication and lexical diversity) were positively intercorrelated at $r = .34$, and our grammar markers (MLU, grammatical complexity, and verb usage) correlated between .79 and .91. Because the grammar markers' correlations were above .60, we computed a grammar composite score by summing the z-scores for grammatical complexity, MLU, and verb usage, in line with our pre-registration. All our speech markers correlated positively with the mothers' reading scores from the Wide-Range Achievement Test (WRAT-3 ⁴⁶). Mothers' complexity of speech and score on WRAT-3 have been shown elsewhere ¹⁰⁰ to be highly correlated, which aligns with our findings here.

SES indicators

Mothers' educational attainment

The highest educational qualification of mothers when the children were aged 5 were coded into 5 groups (1= No Qualifications; 2 = Certificate of Secondary Education (2-5)/General Certificate of Secondary Education (D-G), i.e. below average school leaving qualification; 3= Certificate of Secondary Education (1)/O Level (A-C)/ General Certificate of Secondary Education (A-C), i.e. above average school qualification; 4= Advanced Level/Scholarship Level; 5=Higher National Certificate/Degree i.e., College Level Qualification; 6= Higher National Diploma/Postgraduate degree).

Mothers' occupation

The mothers' current (or most recent) occupation when the children were aged 5 were coded using the Office of Population Censuses and Surveys (1991) Standard Occupational Classification. Occupational groups are arranged into six social classes (1=professional occupations; 2=managerial and technical occupations; 3=skilled occupations (non-manual); 4=skilled occupations (manual); 5=partly skilled occupations; 6=unskilled occupations).

Household income

Mothers reported the household's total income from all sources before tax in the previous 12 months when the children were aged 5. Families were then coded into 16 income brackets, the lowest bracket representing an annual income of £10,000 or less and the highest representing an annual income of £40,000 or more.

Cognition and literacy

Verbal and non-verbal ability

A short form of the Wechsler Preschool and Primary Scale of Intelligence-Revised ⁴⁴ was used to assess the cognitive ability of young children at age 5 years. Two subtests were administered: a score on the Vocabulary Subtest taken as a measure of verbal ability and a score on the Block Design subtest taken as a measure of non-verbal ability. All raw scores were converted to age-based standard scores.

Reading score

Children completed The Sight Word Efficiency (SWE) subtest of the Test of Word Reading Efficiency (TOWRE), which measures a child's ability to pronounce printed words accurately and fluently ⁴³. The SWE was administered at ages 7 and 10 years to assess the number of real printed words that children can identify accurately within 45 seconds. All raw scores were converted to age-based standard scores.

Educational achievement

Teachers rated the children's performance in English and Mathematics when the twins were aged 7, 10, and 12 years. Teachers compared the target child's performance to that of a typical student of the same age on a 5-point Likert scale (1=Far Below, 2=Somewhat Below, 3=Average, 4=Somewhat Above, 5=Far Above). We initially fitted our latent growth curve models separately for English and Mathematics, but because the pattern of results was interchangeable across subjects, we reported models with summed performance scores in our main manuscript (independent models for English and Mathematics are reported in Supplementary Table 8).

Acknowledgements

We are grateful to the study mothers and fathers, the twins, and the twins' teachers for their participation. Our thanks to Professor Louise Arseneault, a former principal investigator of the E-Risk study, and to the E-Risk team for their dedication, hard work, and insights.

Funding

This project was funded by the Nuffield Foundation [EDO/FR-000023238]. The Environmental Risk (E-Risk) Study is funded by grants from the Medical Research Council (MRC) [G1002190, MR/X010791/1]. Additional support was provided by the National Institute of Child Health and Human Development [HD077482] and the Jacobs Foundation. Sophie von Stumm was a Jacobs Foundation CRISP Fellow (2022-2027) and a Paris Institute of Advanced Study Fellow (2023-2024) during the writing of this manuscript. Helen L. Fisher was supported by the Economic and Social Research Council (ESRC) Centre for Society and Mental Health at King's College London [ES/S012567/1]. The views expressed are those of the authors and not necessarily those of the ESRC, Jacobs Foundation, MRC, Nuffield Foundation, or King's College London.

References

1. Duncan, G. J. *et al.* School readiness and later achievement. *Dev. Psychol.* **43**, 1428–1446 (2007).
2. Entwisle, D. R., Alexander, K. L. & Olson, L. S. First Grade and Educational Attainment by Age 22: A New Story. *Am. J. Sociol.* **110**, 1458–1502 (2005).
3. Starr, A., Haider, Z. F. & von Stumm, S. Do school grades matter for growing up? Testing the predictive validity of school performance for outcomes in emerging adulthood. *Dev. Psychol.* (2024) doi:10.1037/dev0001548.
4. Bourdieu, P. *The Forms of Capital*. vol. 1 (1986).
5. Carlson, R. & McChesney, C. Income Sustainability Through Educational Attainment. SSRN Scholarly Paper at <https://papers.ssrn.com/abstract=2676859> (2015).
6. Crawford, C. & Greaves, E. *Socio-Economic, Ethnic and Gender Differences in HE Participation*. (BIS Research Papers 186). UK Department for Business, Innovation & Skills: London, UK. <https://www.gov.uk/government/publications/higher-education-participation-socio-economic-ethnic-and-gender-differences> (2015).
7. Noble, C., Medin, D., Quail, Z., Young, C. & Carter, M. How Does Participation in Formal Education or Learning for Older People Affect Wellbeing and Cognition? A Systematic Literature Review and Meta-Analysis. *Gerontol. Geriatr. Med.* **7**, 2333721420986027 (2021).
8. Furnée, C. A., Groot, W. & van den Brink, H. M. The health effects of education: a meta-analysis. *Eur. J. Public Health* **18**, 417–421 (2008).
9. Balaj, M. *et al.* Effects of education on adult mortality: a global systematic review and meta-analysis. *Lancet Public Health* **9**, e155–e165 (2024).
10. Magnuson, K., Duncan, G. J., Lee, K. T. H. & Metzger, M. W. Early School Adjustment and Educational Attainment. *Am. Educ. Res. J.* **53**, 1198–1228 (2016).
11. von Stumm, S., Cave, S. N. & Wakeling, P. Persistent association between family socioeconomic status and primary school performance in Britain over 95 years. *Npj Sci. Learn.* **7**, (2022).
12. Hoff, E. Interpreting the Early Language Trajectories of Children from Low SES and Language Minority Homes: Implications for Closing Achievement Gaps. *Dev. Psychol.* **49**, 4–14 (2013).
13. Kidd, E., Donnelly, S. & Christiansen, M. H. Individual Differences in Language Acquisition and Processing. *Trends Cogn. Sci.* **22**, 154–169 (2018).

14. Rowe, M. L. Understanding Socioeconomic Differences in Parents' Speech to Children. *Child Dev. Perspect.* **12**, 122–127 (2018).
15. Schwab, J. F. & Lew-Williams, C. Language learning, socioeconomic status, and child-directed speech. *WIREs Cogn. Sci.* **7**, 264–275 (2016).
16. Rowe, M. L. Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill*. *J. Child Lang.* **35**, 185–205 (2008).
17. Hoff, E. The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development Via Maternal Speech. *Child Dev.* **74**, 1368–1378 (2003).
18. Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J. & Hedges, L. V. Sources of variability in children's language growth. *Cognit. Psychol.* **61**, 343–365 (2010).
19. Rowe, M. L. A Longitudinal Investigation of the Role of Quantity and Quality of Child-Directed Speech in Vocabulary Development. *Child Dev.* **83**, 1762–1774 (2012).
20. Foursha-Stevenson, C., Schembri, T., Nicoladis, E. & Eriksen, C. The Influence of Child-Directed Speech on Word Learning and Comprehension. *J. Psycholinguist. Res.* **46**, 329–343 (2017).
21. Durham, R., Hammer, C., Tomblin, J. & Catts, H. Kindergarten oral language skill: A key variable in the intergenerational transmission of socioeconomic status. *Res. Soc. Stratif. Mobil.* **25**, 294–305 (2007).
22. Pace, A., Alper, R., Burchinal, M. R., Golinkoff, R. M. & Hirsh-Pasek, K. Measuring success: Within and cross-domain predictors of academic and social trajectories in elementary school. *Early Child. Res. Q.* **46**, 112–125 (2019).
23. Lippi-Green, R. *English with an Accent: Language, Ideology and Discrimination in the United States*. (Routledge, London, 2011). doi:10.4324/9780203348802.
24. Hart, B. & Risley, T. R. *Meaningful Differences in the Everyday Experience of Young American Children*. (P.H. Brookes, 1995).
25. van der Veen, C. & van Oers, B. Advances in research on classroom dialogue: learning outcomes and assessments. *Learn. Instr.* **48**, 1–4 (2017).
26. Mercer, N., Wegerif, R. & Dawes, L. Children's Talk and the Development of Reasoning in the Classroom. *Br. Educ. Res. J.* **25**, 95–111 (1999).
27. Duff, F. J., Reen, G., Plunkett, K. & Nation, K. Do infant vocabulary skills predict school-age language and literacy outcomes? *J. Child Psychol. Psychiatry* **56**, 848–856 (2015).

28. Lareau, A. & Weininger, E. B. Cultural capital in educational research: A critical assessment. *Theory Soc.* **32**, 567–606 (2003).
29. Bourdieu, P. & Passeron, J.-C. *Reproduction in Education, Society and Culture*, 2nd Ed. xxvi, 254 (Sage Publications, Inc, Thousand Oaks, CA, US, 1990).
30. Kircher, R. & Fox, S. Multicultural London English and its speakers: a corpus-informed discourse study of standard language ideology and social stereotypes. *J. Multiling. Multicult. Dev.* **42**, 792–810 (2021).
31. Spencer, S., Clegg, J. & Stackhouse, J. Language, social class and education: listening to adolescents' perceptions. *Lang. Educ.* **27**, 129–143 (2013).
32. Caspi, A. *et al.* Maternal expressed emotion predicts children's antisocial behavior problems: using monozygotic-twin differences to identify environmental effects on behavioral development. *Dev. Psychol.* **40**, 149–161 (2004).
33. Moffitt, T. E. & E-Risk Study Team. Teen-aged mothers in contemporary Britain. *J. Child Psychol. Psychiatry* **43**, 727–742 (2002).
34. Odgers, C. L., Caspi, A., Bates, C. J., Sampson, R. J. & Moffitt, T. E. Systematic social observation of children's neighborhoods using Google Street View: a reliable and cost-effective method. *J. Child Psychol. Psychiatry* **53**, 1009–1017 (2012).
35. Gustavson, K., von Soest, T., Karevold, E. & Røysamb, E. Attrition and generalizability in longitudinal studies: findings from a 15-year population-based study and a Monte Carlo simulation study. *BMC Public Health* **12**, 918 (2012).
36. Weizman, Z. O. & Snow, C. E. Lexical output as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Dev. Psychol.* **37**, 265–279 (2001).
37. McCarthy, P. M. & Jarvis, S. MTL-D, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behav. Res. Methods* **42**, 381–392 (2010).
38. Hoff-Ginsberg, E. Maternal speech and the child's development of syntax: a further look. *J. Child Lang.* **17**, 85–99 (1990).
39. Mueller, K. D. *et al.* The Latent Structure and Test–Retest Stability of Connected Language Measures in the Wisconsin Registry for Alzheimer's Prevention (WRAP). *Arch. Clin. Neuropsychol.* **33**, 993–1005 (2018).

40. Preacher, K. J. Latent Growth Curve Models. in *The Reviewer's Guide to Quantitative Methods in the Social Sciences* (Routledge, 2018).
41. Rosseel, Y. lavaan: An R Package for Structural Equation Modeling. *J. Stat. Softw.* **48**, 1–36 (2012).
42. Graham, J. W. Missing data analysis: Making it work in the real world. *Annu. Rev. Psychol.* **60**, 549–576 (2009).
43. Torgesen, J. K., Rashotte, C. A. & Wagner, R. K. *TOWRE: Test of Word Reading Efficiency*. (Pro-ed Austin, TX, 1999).
44. Wechsler, D. Wechsler preschool and primary scale of intelligence--revised. *J. Sch. Psychol.* (1989).
45. Duncan, G. J., Daly, M. C., McDonough, P. & Williams, D. R. Optimal Indicators of Socioeconomic Status for Health Research. *Am. J. Public Health* **92**, 1151–1157 (2002).
46. Snelbaker, A. J., Wilkinson, G. S., Robertson, G. J. & Glutting, J. J. Wide Range Achievement Test 3 (wrat3). in *Understanding Psychological Assessment* (eds. Dorfman, W. I. & Hersen, M.) 259–274 (Springer US, Boston, MA, 2001). doi:10.1007/978-1-4615-1185-4_13.
47. Baron, R. M. & Kenny, D. A. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* **51**, 1173–1182 (1986).
48. Pace, A., Luo, R., Hirsh-Pasek, K. & Golinkoff, R. M. Identifying Pathways Between Socioeconomic Status and Language Development. *Annu. Rev. Linguist.* **3**, 285–308 (2017).
49. Adamson, L. B., Kaiser, A. P., Tamis-LaMonda, C. S., Owen, M. T. & Dimitrova, N. The developmental landscape of early parent-focused language intervention. *Early Child. Res. Q.* **50**, 59–67 (2020).
50. Ferjan Ramírez, N., Lytle, S. R. & Kuhl, P. K. Parent coaching increases conversational turns and advances infant language development. *Proc. Natl. Acad. Sci.* **117**, 3484–3491 (2020).
51. American Psychological Association. Education and Socioeconomic Status Factsheet. <https://www.apa.org/pi/ses/resources/publications/education> (2017).
52. Oakes, J. M. & Rossi, P. H. The measurement of SES in health research: current practice and steps toward a new approach. *Soc. Sci. Med.* **1982** **56**, 769–784 (2003).
53. Gennetian, L. A. *et al.* Effects of a monthly unconditional cash transfer starting at birth on family investments among US families with low income. *Nat. Hum. Behav.* **8**, 1514–1529 (2024).

54. d'Apice, K. & Stumm, S. von. Does Age Moderate the Influence of Early Life Language Experiences? A Naturalistic Home Observation Study. Preprint at <https://doi.org/10.31234/osf.io/jr4by> (2019).
55. Huttenlocher, J., Vasilyeva, M., Cymerman, E. & Levine, S. Language input and child syntax. *Cognit. Psychol.* **45**, 337–374 (2002).
56. von Stumm, S., O'Reilly, J. & d'Apice, K. Predicting developmental outcomes in middle childhood from early life language and parenting experiences. *Br. J. Dev. Psychol.* **40**, 487–503 (2022).
57. Sperry, D. E., Sperry, L. L. & Miller, P. J. Language Does Matter: But There is More to Language Than Vocabulary and Directed Speech. *Child Dev.* **90**, 993–997 (2019).
58. Dolean, D., Melby-Lervåg, M., Tincas, I., Damsa, C. & Lervåg, A. Achievement gap: Socioeconomic status affects reading development beyond language and cognition in children facing poverty. *Learn. Instr.* **63**, 101218 (2019).
59. Hart, S. A., Soden, B., Johnson, W., Schatschneider, C. & Taylor, J. Expanding the environment: gene × school-level SES interaction on reading comprehension. *J. Child Psychol. Psychiatry* **54**, 1047–1055 (2013).
60. Caro, D. H., McDonald, J. T. & Willms, J. D. Socio-economic status and academic achievement trajectories from childhood to adolescence. *Can. J. Educ. Can. L'éducation* **32**, 558–590 (2009).
61. Egan-Dailey, S. *et al.* Child-directed speech in a large sample of U.S. mothers with low income. *Child Dev.* (2024) doi:10.1111/cdev.14139.
62. Patel, D., Krems, J. A., Stout, M. E., Byrd-Craven, J. & Hawkins, M. A. W. Parents of Children With High Weight Are Viewed as Responsible for Child Weight and Thus Stigmatized. *Psychol. Sci.* **34**, 35–46 (2023).
63. Adair, J. K., Colegrove, K. S.-S. & McManus, M. E. How the Word Gap Argument Negatively Impacts Young Children of Latinx Immigrants' Conceptualizations of Learning. *Harv. Educ. Rev.* **87**, 309–334 (2017).
64. Marmot, M. The health gap: the challenge of an unequal world. *The Lancet* **386**, 2442–2444 (2015).
65. Kelly, Y., Sacker, A., Del Bono, E., Francesconi, M. & Marmot, M. What role for the home learning environment and parenting in reducing the socioeconomic gradient in child development? Findings from the Millennium Cohort Study. *Arch. Dis. Child.* **96**, 832–837 (2011).
66. Bergelson, E. *et al.* Everyday language input and production in 1,001 children from six continents. *Proc. Natl. Acad. Sci.* **120**, e2300671120 (2023).

67. Anderson, N. J., Graham, S. A., Prime, H., Jenkins, J. M. & Madigan, S. Linking Quality and Quantity of Parental Linguistic Input to Child Language Skills: A Meta-Analysis. *Child Dev.* **92**, 484–501 (2021).
68. Dailey, S. & Bergelson, E. Language input to infants of different socioeconomic statuses: A quantitative meta-analysis. *Dev. Sci.* **25**, e13192 (2022).
69. Hoff-Ginsberg, E. The relation of birth order and socioeconomic status to children's language experience and language development. *Appl. Psycholinguist.* **19**, 603–629 (1998).
70. Curtin, M. Dialect: Social class. *Int. Encycl. Linguist. Anthropol.* 1–9 (2020).
71. Kraus, M. W., Torrez, B., Park, J. W. & Ghayebi, F. Evidence for the reproduction of social class in brief speech. *Proc. Natl. Acad. Sci.* **116**, 22998–23003 (2019).
72. Ochs, E. & Schieffelin, B. Language Socialization: An Historical Overview. in *Language Socialization* (eds. Duff, P. A. & May, S.) 1–14 (Springer International Publishing, Cham, 2017). doi:10.1007/978-3-319-02327-4_1-2.
73. Hoff, E. & Naigles, L. How Children Use Input to Acquire a Lexicon. *Child Dev.* **73**, 418–433 (2002).
74. Schleppegrell, M. J. Linguistic Features of the Language of Schooling. *Linguist. Educ.* **12**, 431–459 (2001).
75. Uccelli, P., Galloway, E. P., Barr, C. D., Meneses, A. & Dobbs, C. L. Beyond Vocabulary: Exploring Cross-Disciplinary Academic-Language Proficiency and Its Association With Reading Comprehension. *Read. Res. Q.* **50**, 337–356 (2015).
76. Nagy, W. & Townsend, D. Words as Tools: Learning Academic Vocabulary as Language Acquisition. *Read. Res. Q.* **47**, 91–108 (2012).
77. Uccelli, P. The Language Demands of Analytical Reading and Writing at School. *Writ. Commun.* **40**, 518–554 (2023).
78. Bakopoulou, I. *Research Report - Empowering Educational Professionals to Create Strong Language Provision in Early and Primary School Settings. Supporting Spoken Language in the Classroom (SSLiC): London, UK.* <https://www.ucl.ac.uk/ioe/departments-and-centres/centres/centre-inclusive-education/supporting-spoken-language-classroom-sslic> (2023).
79. Dickinson, D. K., Golinkoff, R. M. & Hirsh-Pasek, K. Speaking Out for Language: Why Language Is Central to Reading Development. *Educ. Res.* **39**, 305–310 (2010).

80. Magalhães, S., Carneiro, L., Limpo, T. & Filipe, M. Executive functions predict literacy and mathematics achievements: The unique contribution of cognitive flexibility in grades 2, 4, and 6. *Child Neuropsychol.* **26**, 934–952 (2020).
81. Ratcliff, N. J. *et al.* Causes of and Solutions to the Achievement Gap: Teachers' Perceptions. *Teach. Educ. J.* **9**, 97–111 (2016).
82. Gottfried, M. A. Evaluating the Relationship Between Student Attendance and Achievement in Urban Elementary and Middle Schools: An Instrumental Variables Approach. *Am. Educ. Res. J.* **47**, 434–465 (2010).
83. Beck, I. L., McKeown, M. G. & Kucan, L. *Bringing Words to Life: Robust Vocabulary Instruction*. (Guilford Press, 2013).
84. Oslund, E. L., Clemens, N. H., Simmons, D. C., Smith, S. L. & Simmons, L. E. How vocabulary knowledge of middle-school students from low socioeconomic backgrounds influences comprehension processes and outcomes. *Learn. Individ. Differ.* **45**, 159–165 (2016).
85. Vadivel, B., Alam, S., Nikpoo, I. & Ajanil, B. The Impact of Low Socioeconomic Background on a Child's Educational Achievements. *Educ. Res. Int.* **2023**, 6565088 (2023).
86. Nielsen, M., Haun, D., Kärtner, J. & Legare, C. H. The persistent sampling bias in developmental psychology: A call to action. *J. Exp. Child Psychol.* **162**, 31–38 (2017).
87. Hoff-Ginsberg, E. Mother-child conversation in different social classes and communicative settings. *Child Dev.* **62**, 782–796 (1991).
88. Chen, J., Justice, L. M., Tambyraja, S. R. & Sawyer, B. Exploring the mechanism through which peer effects operate in preschool classrooms to influence language growth. *Early Child. Res. Q.* **53**, 1–10 (2020).
89. Pancsofar, N. & Vernon-Feagans, L. Mother and father language input to young children: Contributions to later language development. *J. Appl. Dev. Psychol.* **27**, 571–587 (2006).
90. Barnes, E. M. & Dickinson, D. K. The Relationship of Head Start Teachers' Academic Language Use and Children's Receptive Vocabulary. *Early Educ. Dev.* **28**, 794–809 (2017).
91. Cabral-Gouveia, C., Menezes, I. & Neves, T. Educational strategies to reduce the achievement gap: a systematic review. *Front. Educ.* **8**, (2023).
92. Cushing, I. 'Say it like the Queen': the standard language ideology and language policy making in English primary schools. *Lang. Cult. Curric.* **34**, 321–336 (2021).

93. Trouton, A., Spinath, F. M. & Plomin, R. Twins early development study (TEDS): a multivariate, longitudinal genetic investigation of language, cognition and behavior problems in childhood. *Twin Res. Off. J. Int. Soc. Twin Stud.* **5**, 444–448 (2002).
94. Odgers, C. L. *et al.* Supportive parenting mediates neighborhood socioeconomic disparities in children's antisocial behavior from ages 5 to 12. *Dev. Psychopathol.* **24**, 705–721 (2012).
95. MacWhinney, B. The CHILDES Project. *Comput. Linguist.* **26**, 657–657 (2000).
96. Chall, J. S. & Dale, Edgar. *Readability Revisited : The New Dale-Chall Readability Formula*. (Brookline Books, 1995).
97. Malvern, D., Richards, B., Chipere, N. & Durán, P. Lexical Diversity and Lexical Sophistication in First Language Writing. in *Lexical Diversity and Language Development: Quantification and Assessment* (eds. Malvern, D., Richards, B., Chipere, N. & Durán, P.) 152–176 (Palgrave Macmillan UK, London, 2004). doi:10.1057/9780230511804_9.
98. McKee, G., Malvern, D. & Richards, B. Measuring vocabulary diversity using dedicated software. *Lit. Linguist. Comput.* **15**, 323–338 (2000).
99. Mueller, K. D., Kosciak, R. L., Hermann, B. P., Johnson, S. C. & Turkstra, L. S. Declines in Connected Language Are Associated with Very Early Mild Cognitive Impairment: Results from the Wisconsin Registry for Alzheimer's Prevention. *Front. Aging Neurosci.* **9**, (2018).
100. Golden, C. J. & Bennett, R. *Clinical Integration of Neuropsychological Test Results*. (CRC Press, 2024).