Institute of Education

The intergenerational transmission of vocabulary

By Alice Sullivan, Vanessa Moulton, Emla Fitzsimons

Centre for Longitudinal Studies
Working paper 2017/14

CENTRE FOR LONGITUDINAL STUDIES

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Contact the author
Emla Fitzsimons
Professor of Economics and Director of the Millennium Cohort Study
Centre for Longitudinal Studies
UCL
E.Fitzsimons@ucl.ac.uk

This working paper was first published in November 2017
by the Centre for Longitudinal Studies,
UCL Institute of Education
University College London
20 Bedford Way
London WC1H 0AL
www.cls.ioe.ac.uk

The Centre for Longitudinal Studies (CLS) is an Economic and Social Research Council (ESRC) Resource Centre based at the UCL Institution of Education (IOE). It manages four internationally-renowned cohort studies: the 1958 National Child Development Study, the 1970 British Cohort Study, Next Steps, and the Millennium Cohort Study. For more information, visit www.cls.ioe.ac.uk.

This document is available in alternative formats.
Please contact the Centre for Longitudinal Studies.
tel: +44 (0)20 7612 6875
email: clsfeedback@ioe.ac.uk

Contents

Abstract

Abstract: This paper examines the relationship between parents' and children's vocabulary scores for a nationally representative birth cohort born in the UK - the Millennium Cohort Study (MCS). We investigate both socio-economic and ethnic differentials in children's vocabulary scores, and the role of differences in parents' vocabulary scores in accounting for these. We find large vocabulary gaps between highly educated and less educated parents, and between ethnic groups. Nevertheless, socio-economic and ethnic gaps in vocabulary scores are far wider among the parents than among their children. Parental vocabulary is a powerful mediator of inequalities in offspring's vocabulary scores at age 14, and also a powerful driver of change in language skills between the ages of five and 14. Once we account for parental vocabulary, no ethnic minority group of young people has a negative 'vocabulary gap' compared to whites.

Keywords: vocabulary, education, ethnicity, intergenerational, inequality.

Introduction

Social class differences in language use have been central to some of the major sociological theories regarding the intergenerational transmission of educational advantage. For Bourdieu, sophisticated language use is central to educational and social reproduction (Bourdieu, Passeron and Saint-Martin 1994). Despite this, measures of linguistic attainment have rarely been used in empirical operationalisations of Bourdieu's theory of cultural reproduction (Sullivan 2001). Bernstein also placed substantial weight on the differences in language use by middle and working class children, arguing that this affected their ability to succeed at school (Bernstein 1971; Bernstein 1973; Bernstein 1975). Language knowledge is clearly an important prerequisite for school learning, and language difficulties have been linked to a range of adverse outcomes (Law et al. 2009). Nevertheless, vocabulary and language have latterly been relatively neglected by sociologists, and most of the evidence in this field has come from small-scale studies by psychologists. It may be that sociologists have shied away from the study of language development due to a perception that talking about language and disadvantage may lead to accusations of a deficit approach (Labov 1972).

A number of small scale studies have provided insight on social class and ethnic differences in language skills and patterns. Hart and Risley's (1995) observational study of 42 families in one US college town found strong social class and black-white differences in the range of vocabulary used by parents when talking to their children. Their headline finding that 'upper class' children had been exposed to 30 million more words than 'welfare children' had by age three (Hart and Risley 2003) has been enormously influential, despite the drawback of a small and unrepresentative sample. (To be clear, the 30 million figure does not refer to unique words, but the total barrage of speech to which the children were exposed, including repetition). Subsequent small observational studies have examined socio-economic differences in factors such as the properties of maternal speech (Hoff 2003), the extent of child-directed speech (Rowe 2008) and the use of gestures (Rowe and Goldin-Meadow 2009). We are aware of only one previous large quantitative study which assesses the role of parental vocabulary in shaping class and race differences in children's vocabularies (Farkas and Beron 2004). That paper uses pooled survey waves of the Children of the NLSY79 (CNLSY), a survey of the children of female members of the National Longitudinal Survey of Youth. Farkas and Beron are primarily concerned with examining the longitudinal

growth of class and race vocabulary gaps up to age 13, but, for a subsample of 2969 individuals, they are also able to examine the role of maternal vocabulary in mediating SES and racial vocabulary gaps, and find only a modest mediating effect.

Socio-economic differentials in both verbal and general cognitive attainment emerge early in life, and widen during the pre-school and school years (Becker 2011; Byford, Kuh and Richards 2011; Douglas 1964; Duncan et al. 1998; Farkas and Beron 2004; Feinstein 2003; Fogelman and Goldstein 1976; Law, McBean and Rush 2011; Sullivan and Brown 2015 ; Sullivan, Ketende and Joshi 2013). The vast majority of the literature examining the relationship between parental and child vocabulary is focused on the early years, possibly because it is during this period that the greatest challenges are met (e.g. (Cartmill et al. 2013; Fernald, Marchman and Weisleder 2013; Rowe 2012). In a small scale experiment, Fernald et al (2013) is the first study to find that SES differences in vocabulary and language processing efficiency emerge as early as 18 months, with a 6-month gap emerging between higher- and lower-SES toddlers by 24 months. However, social class differentials in vocabulary continue to grow during adolescence and into mid-life (Sullivan and Brown 2015 ; Sullivan and Brown 2015). The role of differences in parenting practices in explaining the emergence and growth of cognitive gaps has been extensively explored, but the role of parental cognitive resources, including language ability, is relatively neglected, perhaps due to a lack of data on parental cognition in most large-scale datasets.

Compared to early childhood, there is much less research on the influence of the parent-child relationship on adolescent learning. However, vocabulary acquisition in late childhood and adolescence undergoes major quantitative and qualitative changes. Quantitatively, individuals learn a substantial number of words during this period, with evidence that the rate of acquisition of meaning recognition vocabulary during late primary and secondary school years is between 3,000 and 5,400 words per year (Landauer and Dumais 1997). Vocabulary is acquired via indirect or incidental exposure to language materials including school activities, books, the Internet, cinema, TV and radio, and interaction with peers (Gobet 2015; Messer, Dockrell and Murphy 2004). Qualitatively, vocabulary becomes more sophisticated (Berman 2007) as young people start to use longer, more formal, and less common words. A further development occurring around the beginning of secondary school is that teenagers become more sensitive to linguistic registers and can alternate between casual conversation and more formal language. This evolution continues well into adulthood.

Findings regarding ethnic gaps in both cognitive attainment at a given point in time and progress over time vary widely according to the minority groups considered and the national context. In the US context, black-white test score gaps form in early childhood and widen during the school years (Jencks and Phillips 2011; Quinn 2015). In the UK, children of Pakistani and Bangladeshi ethnic backgrounds make greater progress during early schooling than whites (Sullivan, Ketende and Joshi 2013), and white and black Caribbean working class students are generally the lowest achievers at school (Strand 2014). Speaking a language other than English at home is likely to be a relevant factor for some first generation immigrant ethnic groups, but not for established minorities and those from English-speaking countries of origin, such as the Caribbean.

As language skills are associated with educational attainment, educational homogamy is likely to imply that individuals with large vocabularies settle with partners who also have large vocabularies. The children of high-vocabulary unions are likely to benefit both from the breadth of vocabulary in their interactions with each parent and from exposure to the conversations that their parents have with one another. This is likely to promote oral vocabulary development, which in turn feeds in to later reading ability (Beron and Farkas 2004). The tendency for people to marry or partner with others of similar educational status

(known as homogamy, or assortative mating) is well established (Blackwell 1998; Blossfeld and Timm 2003; Elder Jr 1969; Kalmijn 1998; Reynolds, Baker and Pedersen 2000; Uunk, Ganzeboom and Róbert 1996) and has been implicated in growing resource inequalities between households (Blossfeld and Buchholz 2009).

In this paper we exploit data on the vocabulary scores of both parents and children in a nationally representative UK birth cohort study. A unique feature of the study is that mothers, partners and children took an equivalent vocabulary test when the children were aged 14. This allows us to build on the existing evidence base in a number of ways. First of all, we are able to establish the vocabulary gaps that exist for both parents and children according to social class and ethnic group using a nationally representative birth cohort study. Second, we address the extent to which socio-economic and ethnic gaps in children's scores at age 14 are driven by differences in the parents' scores. Third, we assess whether parental vocabulary is associated with a growing language gap for children up to age 14.

Theoretical background and research questions

From the perspective of cultural reproduction theory, the ability to understand and use 'educated' language is a vital part of the advantage which is transmitted by high status parents (Bourdieu 1977). In this sense, language can be seen as part of the cultural capital that is transmitted within the home. Of course, the concept of cultural capital has been operationalised in diverse ways. A useful distinction has been drawn between 'status-seeking' and 'information processing' forms of cultural capital (Ganzeboom 1982). Information processing cultural capital leads to the development of knowledge and skills which are rewarded in the education system (Sullivan 2002). Status-seeking cultural capital is rewarded via teacher bias rather than improved skills (Farkas et al. 1990; Jæger and Møllegaard 2017). Within this framework, activities connected with books and reading are seen as 'information processing', whereas other activities such as playing a musical instrument, attending art galleries and museums, etc. are seen as 'status seeking', though we prefer the terms 'literary' and 'non-literary' cultural capital. Studies which have separated the two have found that literary cultural capital has more influence on educational attainment (De Graaf, De Graaf and Kraaykamp 2000; Jæger 2011; Sullivan 2001). However, studies of cultural reproduction have neglected the role of parental language skills. This means we do not know to what extent parental language skills are transmitted to the child, or how important parental language skills are in explaining SES gaps in children's language skills.

Psychologists have also stressed the importance of the home literacy environment to children's language learning (Melhuish et al. 2008; Waldfogel and Washbrook 2011). Three aspects of parenting have been highlighted as central to children's early language and learning (Rodriguez et al. 2009): (1) frequency of children's participation in routine *learning activities* (e.g., shared book reading, storytelling); (2) the *quality of caregiver-child engagements* (e.g., parents' cognitive stimulation and sensitivity/responsiveness); and (3) the provision of age-appropriate *learning materials* (e.g., books and toys). Studies have found substantial socio-economic differentials in these parental inputs (Bassok et al. 2016). However, studies assessing the role of the home literacy environment have not accounted for the role of parental language skills. This is important, because it is likely both that parents who have strong language skills will be most comfortable engaging in activities such as shared reading, and also that they may be more effective at engaging their children in these activities than, for example, a parent who struggles with basic literacy skills (Sullivan, Ketende and Joshi 2013).

Regardless of the theoretical perspective applied, empirical findings across the disciplines highlight the importance of the home literary climate. Whether books in the home are termed 'embodied cultural capital' or 'learning materials', they remain a powerful predictor of educational outcomes (Marks, Cresswell and Ainley 2006). Both theoretical perspectives have merit, as books in the home are learning materials, but also reflect the value placed upon books and learning within the family, and represent a cultural display, signalling that the owner is a cultured person. Parental reading to children (Bus, Van Ijzendoorn and Pellegrini 1995) and children's own reading (Stanovich and Cunningham 1998; Sullivan and Brown 2015) are powerful predictors of language learning and general educational outcomes.

In this paper, we attempt to assess the role of parental vocabulary in explaining socioeconomic and ethnic differentials in children's vocabulary, also taking on board the roles of the home literary environment and the child's own cultural practices.

Research questions

1. How strongly correlated are the vocabulary scores of the mother, partner and child?
2. How large are vocabulary gaps according to childhood socio-economic circumstances, ethnic group, and other factors?
3. What is the role of the home literacy culture in predicting child vocabulary and explaining SES gaps?
4. What is the role of the child's own cultural capital in predicting vocabulary and explaining SES gaps? We assess the roles of reading for pleasure (literary cultural capital) and playing an instrument (non-literary cultural capital).
5. How important are the mother's and partner's vocabulary in predicting the child's vocabulary, and does this substantially mediate SES and other effects in the model?
6. Which factors are relevant for progress in verbal scores between the ages of five and 14?

Data and analytical approach

The Millennium Cohort Study (MCS) is a national birth cohort study following the lives of 19,517 children born in the UK in 2000-01 (Connelly and Platt 2014; Joshi and Fitzsimons 2016). There have been six main sweeps of data collection, at ages 9 months and 3, 5, 7, 11 and 14 years. The study is multi-disciplinary, and contains rich measures of childhood socio-economic circumstances, child development and child health. The MCS datasets are freely available to researchers internationally via the UK Data Service (http://ukdataservice.ac.uk). The MCS website provides detailed information and documentation on the study (http://www.cls.ioe.ac.uk/mcs).

Our sample (n = 11,714) comprised all households at MCS6. We used records for only one child per family (singletons and the first-born twin or triplets) to avoid having to account for clustering of children within families. We exploit data provided by the 'main respondent' parent (this is typically the mother, and we refer henceforth in the text to mothers rather than main respondents), the spouse or cohabiting partner (where applicable), and the child themselves, up to age 14.

Our initial, descriptive, analysis first assesses the correlations between the mother, partner and child's vocabulary scores, assessed in 2015, when the cohort member was aged 14. The vocabulary scores were derived from a shortened version of the Applied Psychology Unit (APU) Vocabulary Test, a standardised test produced by the University of Edinburgh (Closs, 1976), and used in previous studies including the 1970 British Cohort Study (BCS70) (Parsons 2014). The APU Vocabulary Test directly examines vocabulary knowledge,

through multiple-choice items in which a stimulus word has to be matched to a synonym from five alternatives. At the start of the test the stimulus words are very easy e.g. 'begin' and become more progressively difficult; the last word in the test is 'pusillanimous'. The final score is the sum of the correct answers, from a total of 20 multiple-choice items. The test words were selected from national newspapers and magazines, with the intention of excluding archaic, over-specialised and little-used words (Levy & Goldstein, 1984). Reliability of the measure has been satisfactory; internal consistency with the 15-16 age range was 0.95 and retest reliability with 13 year olds of 0.83 (Levy & Goldstein, 1984). Concurrent validity of 0.78 was found with 13 year olds on the Mill Hill Vocabulary Test (which measures verbal intelligence and has a 0.75 correlation with the Raven's Progressive Matrices; Raven, 1983).

We present the mean parental and child vocabulary scores according to the other variables to be used in our models. This is followed by a series of linear regression models, with child vocabulary at age 14 as the outcome, as follows:

Model 1: Socio-economic and demographic factors

Our first model includes essential socio-economic and demographic information, and provides an indication of the magnitude of the associations between these variables and the child's vocabulary scores before any potential mediating factors have been accounted for. This model includes the age (in months), sex and ethnic group of the child, and the region of the UK that the family live in. Parents' education is the highest qualification of either parent. Economic circumstances are captured wave 1 of the survey in 2001-2 (or wave 2 in 2004-5 if not available at wave 1) by parental social class measured on the NS-SEC scale (Goldthorpe and McKnight 2006), home ownership, and log equivalised family income. The number of older and younger siblings is included, as older siblings have been shown to be advantaged both in terms of vocabulary and general cognitive outcomes (Hoff-Ginsberg 1998; Nisbet 1953). Further, mothers differ in their language, engagement and responsiveness toward their first- and later-born children, with input favouring firstborns (Bornstein 2002). Whether English is the main language spoken at home at wave 1 of the survey (or wave 2 where unavailable at wave 1) is included, as this may be related to both parental and child vocabulary scores. Maternal and partner age at the 2015 interview are also included as controls. In addition, the models controlled for single parent household status at wave 6 (in 2015), and if, the mother's partner was present, whether they completed the vocabulary test or not.

Model 2: Model 1 + Home literary climate

Model 2 adds books in the home and the frequency of parental reading to the child at age three.

Model 3: Model 2 + Child's cultural capital

Model 3 adds the child's own reading and playing a musical instrument at age 11. This assesses the roles of literary vs non-literary cultural capital.

Model 4: Model 3 + parental vocabulary

The mother and partner's vocabulary scores are added to this model, to assess the extent to which this accounts for the effects of socio-economic and demographic circumstances and of parental and child cultural capital.

Model 5: Model 4 + age 5 cognitive scores

Several studies have shown that language scores on school entry are a strong predictor of later language acquisition (Duncan et al. 2007). This model includes cognitive scores at the start of schooling (Naming Vocabulary, Pattern Construction, and Picture Similarities from the second edition of the British Ability Scales (Elliott, 1996)), in order to assess the extent to which the effects of the childhood circumstances included in our model are fully captured by cognitive scores at age five, around the start of formal schooling, and the extent to which differentials continue to grow between the ages of five and 14.

We exploit data from birth to age 14, and, as in any longitudinal analysis, the problem of missing data must be addressed (Mostafa and Wiggins 2015). It is well known that list-wise deletion/complete case analysis returns biased estimates, so we use multiple imputation with chained equations (25 imputed datasets) to 'fill-in' values of any missing items in the variables selected for our analysis adopting Schafer's data augmentation approach (Schafer 1997) under the assumption of 'missing at random' (MAR). In order to maximise the plausibility of the MAR assumption we also included a set of auxiliary variables in our imputation model. In this instance MAR implies that that our estimates are valid if 'missingness' is due to variables (auxiliary or substantive) that were included in our models (Little and Rubin 2002). In addition, to take account of disproportionate, stratified clustering in the MCS sample design, models were adjusted for non-response and the MCS survey design (Hansen 2012).

Results

Descriptive results

FIGURES 1, 2, 3

Figures 1, 2 and 3 show the distributions of the scores for the young person, mother and partner. Young people achieved a mean score of seven out of 20 on average, while mothers and partners gained substantially higher scores (11 and 12 respectively).

TABLE 1

Table 1 presents raw (imputed and weighted) mean scores out of 20 in the vocabulary assessment, by respondent type (young person, main respondent (mother) and partner). We also analysed the vocabulary mean scores for mothers with partners (available on request), who had slightly higher vocabulary scores than mothers without partners, but the difference is not statistically significant at the 0.05 level.

We observe stark ethnic differences (based on the young person's ethnic identification) in adult vocabulary scores. The parents of white and ethnically mixed young people had the highest mean scores (between 10.9 and 12.1) – around two and a half times higher than the parents of ethnically Bangladeshi young people, who received the lowest mean scores (between 4.2 and 4.6). Such stark differences by ethnicity are not observed amongst the young people themselves, the ethnic gaps in their scores are far more modest (ranging between a mean score of 6.1 for Pakistanis and 6.9 for ethnically mixed and white young people). Regional differences in adult vocabulary scores are also apparent, with those living in London scoring lowest, reflecting the ethnically diverse population.

There are strong gradients in parental vocabulary scores according to parental education, social class and income. Among main respondents where at least one partner had a university degree, the mean score was 15, compared to 6.5 for households where neither parent had any educational qualification. The education gaps are less marked for the offspring than for the parents. The children of university graduates scored 8.6 versus 5.8 for

children in households with no qualifications. Another way of looking at this is that the vocabulary gap between parents and children widens with parents' education. For families with no parental qualifications, the average mean score for children (5.8) is only around one correct answer less than for mothers (6.6). Essentially this means that both mothers and children got roughly a third of the items right on average. For families where one parent has a higher degree, children scored an average of 8.6 compared to 15.0 for mothers, meaning that the children got less than half of the items correct, while the mothers got three-quarters correct. A similar pattern is observed for social class, household income and home ownership– parental socio-economic vocabulary gaps are larger than those for young people.

Mean adult English vocabulary is considerably lower amongst those whose home language is mixed or non-English, compared to English only (6.6 and 10.7 respectively for main respondents), but the difference among young people is negligible (6.6 versus 6.8).

Turning to indicators of cultural resources, we see that both parents' and children's vocabulary scores are higher in households with higher levels of books in home and more frequent reading to the child. Young people who read frequently have relatively high mean vocabulary scores, whereas the adult scores are less strongly differentiated according to this measure. The vocabulary gap between young people who play a musical instrument and those who don't is small (7.3 vs 6.6).

The young person's vocabulary score varies according to both the mother's and partner's scores. For example, children whose mothers scored in the top quartile achieved mean scores of 8.2 compared to 5.9 for children whose mothers scored in the bottom quartile. A similar pattern is observed according to partner's vocabulary scores.

Table 1 also shows how scores vary by children's earlier cognitive scores, as measured at age five: there is a positive relationship between children's earlier cognitive measures and their vocabulary at age 14, particularly for verbal cognition. Children in the bottom quartile for verbal cognition at age five scored a mean of 5.7 words at age 14, compared to those in the top quartile of cognitive scores at age five, who achieved an average score of 7.9 at age 14.

Table 2 shows a correlation matrix of the young person, main respondent and partner vocabulary scores and the child's early cognitive scores. The main respondent and partner scores are highly correlated, at around 0.5; correlations of around 0.3 are observed between the young person and main/partner. We also see higher correlations between earlier verbal cognition and age 14 vocabulary (0.32) than between early measures of spatial and pictorial reasoning and age 14 vocabulary (0.24 and 0.19 respectively).

Regression results

TABLE 3

Table 3 shows a series of models predicting vocabulary scores at age 14. The outcome variable, parental vocabulary scores, and the child's prior cognitive scores are all standardised to mean=0, standard deviation=1. Parental education is clearly the most powerful predictor in our first model, and is far more strongly linked to the child's vocabulary than social class, income or home ownership. Having an undergraduate (bachelors) university degree or a higher (postgraduate) degree (compared to no qualifications) provides roughly three times the advantage associated with having a parent with a higher managerial or professional occupation (compared to a routine occupation) when both are included in the same model.

Income and home ownership are not significantly associated with vocabulary, taking the other factors in the model into account.

Some ethnic groups are disadvantaged in terms of vocabulary in this model: young people who are identified as ethnically Indian, Pakistani, black African or 'other' have statistically significantly lower scores than whites. There is no difference between boys and girls. As expected, young people with older siblings have a significant disadvantage in vocabulary scores, while the existence and number of younger siblings makes little difference. Speaking a language other than English at home is, perhaps surprisingly, positively linked to child vocabulary, once other factors, notably ethnic group, are factored in. The country of the UK is included, with London split from the rest of England, as educational policy and practice vary across these regions. Living in Scotland or Wales is associated with a disadvantage in vocabulary. We include the ages of the mother, partner and child in the model, as vocabulary is expected to increase with age, especially in the case of the child. Both maternal age (in years) and child age (in months) are positively associated with the child's vocabulary score. The age range of the MCS births extended over a full calendar year. The coefficient associated with a month in age (0.01) can therefore usefully be compared to other coefficients in the model. For example, having a parent with a first degree is associated with five times the vocabulary advantage associated with one year in age.

Model 2 introduces parental cultural resources, in the form of books in the home and reading to the child at age three. Both of these variables are positive predictors of the young person's vocabulary, and books in the home is particularly powerful. The introduction of these variables mediates the parental education and social class effects to some extent.

Model 3 includes the child's own reading for pleasure, and paying a musical instrument. Playing a musical instrument is statistically significant, but, in line with theory, reading has a far more powerful association with child vocabulary. However, these child activities do little to mediate the effects of parental education, social class, and cultural capital.

In model 4, we introduce the mother's and partner's vocabulary scores. Both are strongly independently associated with the child's score, especially the mother's. Parental vocabulary has a powerful mediating effect on parental education, reducing the coefficients for a degree and higher degree by about half, and reducing lower levels of education to statistical insignificance. Social class also becomes non-significant in this model. The effect of the home literary climate is also substantially reduced, but the effect of the child's own cultural activities is unaffected.

All ethnic differences become small and non-significant in this model, with the exception that the Bangladeshi group have a substantial advantage over whites once parental vocabulary is controlled. This suggests that Bangladeshi children have higher vocabulary scores than would be expected based on their parents' scores. The disadvantage associated with living in Wales remains significant.

Our final model includes the child's prior verbal and non-verbal cognitive scores. As expected, these are strong predictors of attainment, particularly the verbal score. This model shows us which factors predict progress between the ages of five and 14. Parental vocabulary remains highly significant. The effects of parental education and books in the home are only slightly reduced, while reading to the child at age three becomes non-significant, its effect being fully captured by the child's cognitive scores at age five. The effect of the child's own reading for pleasure is only slightly reduced, and this variable remains a powerful predictor of vocabulary. The coefficient for girls is negative in this model, suggesting that girls have made less progress than boys between the ages of five and 14. The positive Bangladeshi coefficient becomes

stronger, suggesting that children of Bangladeshi ethnicity have made greater progress than whites with similar characteristics. The negative coefficient for children in Wales remains.

As an indication of effect size, we calculated the meaning of the coefficients in this final model in terms of the raw test scores. Notable coefficients are as follows: a one standard deviation increase in verbal cognition at age five is associated with a 0.5 word increase in mean vocabulary scores (out of 20); a one standard deviation increase in maternal vocabulary is associated with an advantage of 0.4 words; a one standard deviation increase in partner's vocabulary equates to 0.3 words; a higher degree equates to 0.4 words; more than 500 books in the home equates to 0.6 words; Bangladeshi ethnicity equates to 0.6 words; a non-English language at home equates to 0.7 words, and reading for pleasure most days equates to 0.8 words.

This paper has reported on the intergenerational transmission of vocabulary. It is likely that this process has wider implications for educational attainment and social reproduction and mobility. We intend to investigate these in future work.

The raw inequalities that we find in parental vocabulary are startling. For example, parents with an undergraduate degree knew twice as many words on our test as parents with no qualifications. Though of course not directly comparable with Hart and Risley's (1995) study, which was carried out in another place and time, and using different methods, this difference is in line with the order of magnitude of social class differences in vocabulary fount by Hart and Risley's work. However, whereas Hart and Risley found similar social class differences in vocabulary for children as for their parents, we do not. The socio-economic differentials that we found for young people at age 14 were marked, but substantially more modest than those found among their parents. This gives some grounds for optimism, in that socio-economic differentials in vocabulary are not transmitted wholesale from parents to children. Children are exposed to vocabulary, not just from their parents, but from a range of sources including friends, teachers, books, TV and the internet. Some of these wider exposures may mitigate the relationship between parental and child vocabulary.

Our models of children's vocabulary at age 14 show that parental education is far more important than other aspects of socio-economic position in shaping differentials in children's vocabulary scores. The differentials due to parental education were somewhat reduced by accounting for the home literary climate. In contrast, the child's own cultural activities, particularly reading, matter but do not mediate the differential due to parental education. This challenges the traditional cultural reproduction framework, to the extent that the child's own cultural participation appears to have little to do with the reproduction of socio-economic differentials in attainment.

We have shown that parental vocabulary is a vital mediator of differentials in children's vocabulary according to parental education, and parental vocabulary also partly explains the apparent influence of the home literary environment on children's vocabularies. This suggests that the omission of parental vocabulary from most previous models of children's language development, and indeed of their educational development more generally, may have led to a skewed and incomplete understanding of inequalities in children's outcomes. The role of parental vocabulary in accounting for ethnic differentials in children's vocabulary is also powerful – the negative differentials in young people's vocabulary between some minority ethnic groups and whites are fully explained by differences in parental vocabulary. Furthermore, parental vocabulary strongly predicts language progress between the ages of five and 14, suggesting that its influence is not restricted to early childhood development.

Regional differences in children's vocabularies were largely accounted for by demographic factors, but a robust Welsh disadvantage remained, which is consistent with previous international comparisons of academic performance (Jerrim and Shure 2016). This is worthy of further investigation, and may reflect a different educational approach within the devolved Welsh administration (Taylor, Joshi and Wright 2015), and the promotion of the Welsh language within schools. The Welsh finding is interesting given that our analysis suggests that speaking a language other than English in the home is generally positive, in line with other evidence on bilingualism (Marian and Shook 2012).

From a theoretical point of view, our findings support the view that language skills are an important part of the resources that more privileged parents possess and are able, to some degree, to transmit to their children, although this process is far from deterministic. In policy terms, our findings suggest that children whose parents are less educated and those from particular ethnic minority groups may require additional input at school to support the development of a rich vocabulary, and encouraging independent reading is likely to be a useful tool in this regard. In the case of immigrant parents who lack English fluency, support for their English language development is likely to benefit their children.

Bibliography

Bassok, Daphna, Jenna E Finch, RaeHyuck Lee, Sean F Reardon, and Jane Waldfogel. 2016. "Socioeconomic gaps in early childhood experiences: 1998 to 2010." *AERA Open* 2(3):2332858416653924.

Becker, Birgit. 2011. "Social disparities in children's vocabulary in early childhood. Does pre-school education help to close the gap? 1." *The British journal of sociology* 62(1):69-88.

Berman, R.A. 2007. "Developing language knowledge and language use across

adolescence. ." *Hoff, E., Shatz, M. (Eds.), Handbook of Language Development.*

Blackwell Publishing, London.

Bernstein, Basil. 1971. *Class, Codes and Control. Volume 1: Theoretical Studies Towards A Sociology of Language*. London: Routledge and Kegan Paul.

—. 1973. *Class, Codes and Control. Volume 2, Applied Studies towards a Sociology of Language*. London: Routledge and Kegan Paul.

—. 1975. *Class, codes and control, Volume 3: Towards a theory of educational transmissions* London: Routledge and Kegan Paul.

Beron, Kurt J, and George Farkas. 2004. "Oral language and reading success: A structural equation modeling approach." *Structural Equation Modeling* 11(1):110-31.

Blackwell, Debra L. 1998. "Marital homogamy in the United States: The influence of individual and paternal education." *Social Science Research* 27(2):159-88.

Blossfeld, Hans-Peter, and Sandra Buchholz. 2009. "Increasing resource inequality among families in modern societies: The mechanisms of growing educational homogamy, changes in the division of work in the family and the decline of the male breadwinner model." *Journal of Comparative Family Studies*:603-16.

Blossfeld, Hans-Peter, and Andreas Timm. 2003. *Who marries whom?: educational systems as marriage markets in modern societies*: Springer Science & Business Media.

Bornstein, MH. 2002. "Parenting infants." Pp. 3-44 in *Handbook of parenting: Children and parenting*, edited by MH Bornstein. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Bourdieu, P. 1977. "Cultural Reproduction and Social Reproduction." in *Power and ideology in education*, edited by J. Karabel and A. H. Halsey: OUP.

Bourdieu, P., J. -C. Passeron, and M. Saint-Martin. 1994. *Academic Discourse*. Cambridge: Polity Press.

Bus, Adriana G, Marinus H Van Ijzendoorn, and Anthony D Pellegrini. 1995. "Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy." *Review of Educational Research* 65(1):1-21.

Byford, M, D Kuh, and M Richards. 2011. "Parenting practices and intergenerational associations in cognitive ability." *International Journal of Epidemiology*:dyr188.

Cartmill, E. A., B. F. Armstrong, 3rd, L. R. Gleitman, S. Goldin-Meadow, T. N. Medina, and J. C. Trueswell. 2013. "Quality of early parent input predicts child vocabulary 3 years later." *Proc Natl Acad Sci U S A* 110(28):11278-83.

Connelly, Roxanne, and Lucinda Platt. 2014. "Cohort profile: UK millennium Cohort study (MCS)." *International Journal of Epidemiology* 43(6):1719-25.

De Graaf, N.D., P.M. De Graaf, and G. Kraaykamp. 2000. "Parental Cultural Capital and Educational Attainment in the Netherlands: A Refinement of the Cultural Capital Perspective." *Sociology of Education* 73:92-111.

Douglas, J.W.B. 1964. *The Home and the School: A study of ability and attainment in the primary school*. London: MacGibbon and Kee.

Duncan, G., W. Yeung, J. Brooks-Gunn, and J. Smith. 1998. "How much does childhood poverty affect the lifechances of children?" *American Sociological Review* 63:406-23.

Duncan, Greg J, Chantelle J Dowsett, Amy Claessens, Katherine Magnuson, Aletha C Huston, Pamela Klebanov, Linda S Pagani, Leon Feinstein, Mimi Engel, and Jeanne Brooks-Gunn. 2007. "School readiness and later achievement." *Developmental psychology* 43(6):1428.

Elder Jr, Glen H. 1969. "Appearance and education in marriage mobility." *American Sociological Review*:519-33.

Farkas, G., R. P. Grobe, D. Sheehan, and Y. Shuan. 1990. "Cultural Resources and School Success: Gender, Ethnicity and Poverty Groups within an Urban School District." *American Sociological Review* 55:127-42.

Farkas, George, and Kurt Beron. 2004. "The detailed age trajectory of oral vocabulary knowledge: Differences by class and race." *Social Science Research* 33(3):464-97.

Feinstein, L. 2003. "Inequality in the Early Cognitive Development of British Children in the 1970 Cohort." *Economica* 70(1):73-97.

Fernald, A., V. A. Marchman, and A. Weisleder. 2013. "SES differences in language processing skill and vocabulary are evident at 18 months." *Dev Sci* 16(2):234-48.

Fogelman, K.R., and H. Goldstein. 1976. "Social Factors Associated with Changes in Educational Attainment between 7 and 11 Years of Age." *Educational Studies* 2:95-109.

Ganzeboom, H. 1982. "Explaining Differential Participation in High-Cultural Activities - A Confrontation of Information-Processing and Status-Seeking Theories." Pp. 186-205 in *Theoretical Models and Empirical Analyses: Contributions to the Explanation of Individual Actions and Collective Phenomena*, edited by W. Raub. Utrecht: E.S. Publications.

Gobet, F. 2015. "Vocabulary Acquisition." *International encyclopedia of the social & behavioral sciences, 2nd edition, Publisher: Elsevier, Editors: James D. Wright* 25:226-31.

Goldthorpe, J., and A. McKnight. 2006. "The economic basis of social class." in *Mobility and Inequality: Frontiers of research from sociology and economics*, edited by S. Morgan, D.B. Grusky, and G.S. Fields. Stanford: Stanford University Press.

Hansen, K. (Ed.). 2012. *Millennium Cohort Study. First, Second, Third and Fourth Surveys. A Guide to the Datasets.* London: Centre for Longitudinal Studies, Institute of Education.

Hart, Betty, and Todd R Risley. 1995. *Meaningful differences in the everyday experience of young American children*: Paul H Brookes Publishing.

—. 2003. "The early catastrophe: The 30 million word gap by age 3." *American Educator* 27(1):4-9.

Hoff-Ginsberg, Erika. 1998. "The relation of birth order and socioeconomic status to children's language experience and language development." *Applied Psycholinguistics* 19(04):603-29.
Hoff, Erika. 2003. "The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech." *Child development* 74(5):1368-78.
Jæger, Mads Meier. 2011. "Does cultural capital really affect academic achievement? New evidence from combined sibling and panel data." *Sociology of Education* 84(4):281-98.
Jæger, Mads Meier, and Stine Møllegaard. 2017. "Cultural capital, teacher bias, and educational success: New evidence from monozygotic twins." *Social Science Research* 65:130-44.
Jencks, Christopher, and Meredith Phillips. 2011. *The Black-White test score gap*: Brookings Institution Press.
Jerrim, John, and Nikki Shure. 2016. "Achievement of 15 year olds in Wales: PISA 2015 national report." UCL, Institute of Education.
Joshi, HE, and Emla Fitzsimons. 2016. "The UK Millennium Cohort: the making of a multipurpose resource for social science and policy." *Longitudinal and Life Course Studies* 7(4):409-30.
Kalmijn, Matthijs. 1998. "Intermarriage and homogamy: Causes, patterns, trends." *Annual Review of Sociology* 24(1):395-421.
Labov, W. 1972. "The Logic of Nonstandard English." in *Language and Social Context*, edited by P. P. Giglio: Penguin.
Landauer, T. K., and S. T. Dumais. 1997. "A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge." *Psychological Review* 104(2):211-40.
Law, James, Kirsty McBean, and Robert Rush. 2011. "Communication skills in a population of primary school-aged children raised in an area of pronounced social disadvantage." *International journal of language & communication disorders* 46(6):657-64.
Law, James, Robert Rush, Ingrid Schoon, and Samantha Parsons. 2009. "Modeling developmental language difficulties from school entry into adulthood: Literacy, mental health, and employment outcomes." *Journal of Speech, Language, and Hearing Research* 52(6):1401-16.
Little, R. J. A., and D. B. Rubin. 2002. *Statistical Analysis with Missing Data* Chichester: Willey.
Marian, Viorica, and Anthony Shook. 2012. "The cognitive benefits of being bilingual." in *Cerebrum: the Dana forum on brain science*: Dana Foundation.
Marks, Gary N, John Cresswell, and John Ainley. 2006. "Explaining socioeconomic inequalities in student achievement: The role of home and school factors." *Educational Research and Evaluation* 12(02):105-28.
Melhuish, Edward C, Mai B Phan, Kathy Sylva, Pam Sammons, Iram Siraj-Blatchford, and Brenda Taggart. 2008. "Effects of the home learning environment and preschool center experience upon literacy and numeracy development in early primary school." *Journal of Social Issues* 64(1):95-114.
Messer, D., J. E. Dockrell, and N. Murphy. 2004. "Relation between naming and literacy in children with word-finding difficulties." *Journal of Educational Psychology* 96(3):462-70.
Mostafa, Tarek, and R.D. Wiggins. 2015. "The impact of attrition and non-response in birth cohort studies: a need to incorporate missingness strategies." *Longitudinal and Life Course Studies* 6(2):131-46.
Nisbet, J. 1953. "Family Environment and Intelligence." *Eugenics Review* XLV:31-42.
Parsons, S. 2014. "Childhood cognition in the 1970 British Cohort Study." *CLS Data Note*.
Quinn, David M. 2015. "Kindergarten black–white test score gaps: Re-examining the roles of socioeconomic status and school quality with new data." *Sociology of Education* 88(2):120-39.
Reynolds, Chandra A, Laura A Baker, and Nancy L Pedersen. 2000. "Multivariate models of mixed assortment: phenotypic assortment and social homogamy for education and fluid ability." *Behavior genetics* 30(6):455-76.

Rodriguez, E. T., C. S. Tamis-LeMonda, M. E. Spellmann, B. A. Pan, H. Raikes, J. Lugo-Gil, and G. Luze. 2009. "The formative role of home literacy experiences across the first three years of life in children from low-income families." *Journal of Applied Developmental Psychology* 30(6):677-94.

Rowe, M. L. 2012. "A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development." *Child Dev* 83(5):1762-74.

Rowe, Meredith L. 2008. "Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill." *Journal of child language* 35(01):185-205.

Rowe, Meredith L, and Susan Goldin-Meadow. 2009. "Differences in early gesture explain SES disparities in child vocabulary size at school entry." *Science* 323(5916):951-53.

Schafer, J.L. 1997. *Analysis of Incomplete Multivariate Data*. London: Chapman and Hall.

Stanovich, KE, and AE Cunningham. 1998. "What reading does for the mind." *American Education Journal* 22(1-2):8-15.

Strand, Steve. 2014. "Ethnicity, gender, social class and achievement gaps at age 16: Intersectionality and 'Getting it'for the white working class." *Research Papers in Education* 29(2):131-71.

Sullivan, A. 2001. "Cultural Capital and Educational Attainment." *Sociology* 35(4):893-912.

—. 2002. "Bourdieu and Education: How Useful is Bourdieu's Theory for Researchers?" *Netherlands' Journal of Social Sciences* 38(2):144-66.

Sullivan, A., and M. Brown. 2015 "Reading for pleasure and children's progress in vocabulary and mathematics." *British Educational Research Journal* 41(6):971-91.

Sullivan, A., S. Ketende, and H. Joshi. 2013. "Social Class and Inequalities in Early Cognitive Scores." *Sociology* 47(6):1187-206.

Sullivan, Alice, and Matthew Brown. 2015. "Vocabulary from adolescence to middle age." *Longitudinal and Life Course Studies* 6(2):173-89.

Taylor, Chris, Heather Joshi, and Caroline Wright. 2015. "Evaluating the impact of early years educational reform in Wales to age seven: the potential use of the UK Millennium Cohort Study." *Journal of Education Policy* 30(5):688-712.

Uunk, Wilfred JG, Harry BG Ganzeboom, and Péter Róbert. 1996. "Bivariate and multivariate scaled association models. An application to homogamy of social origin and education in Hungary between 1930 and 1979." *Quality & Quantity* 30(3):323-43.

Waldfogel, Jane, and Elizabeth Washbrook. 2011. "Income-related gaps in school readiness in the United States and the United Kingdom." in *Persistence, Privilege and Parenting: The comparative study of intergenerational mobility*, edited by T. Smeeding, R. Erikson, and M. Jantti: Russell Sage Foundation.

Tables

Figure 1: Distribution of young people's vocabulary scores

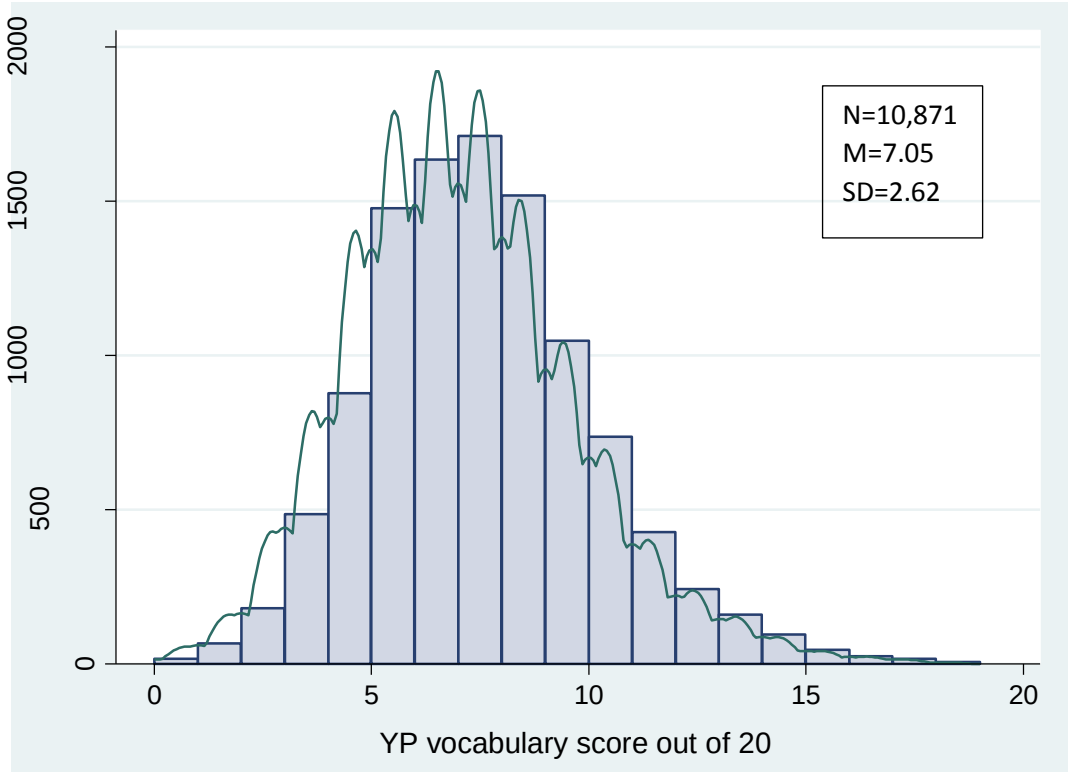

Figure 2: Distribution of mothers' vocabulary scores

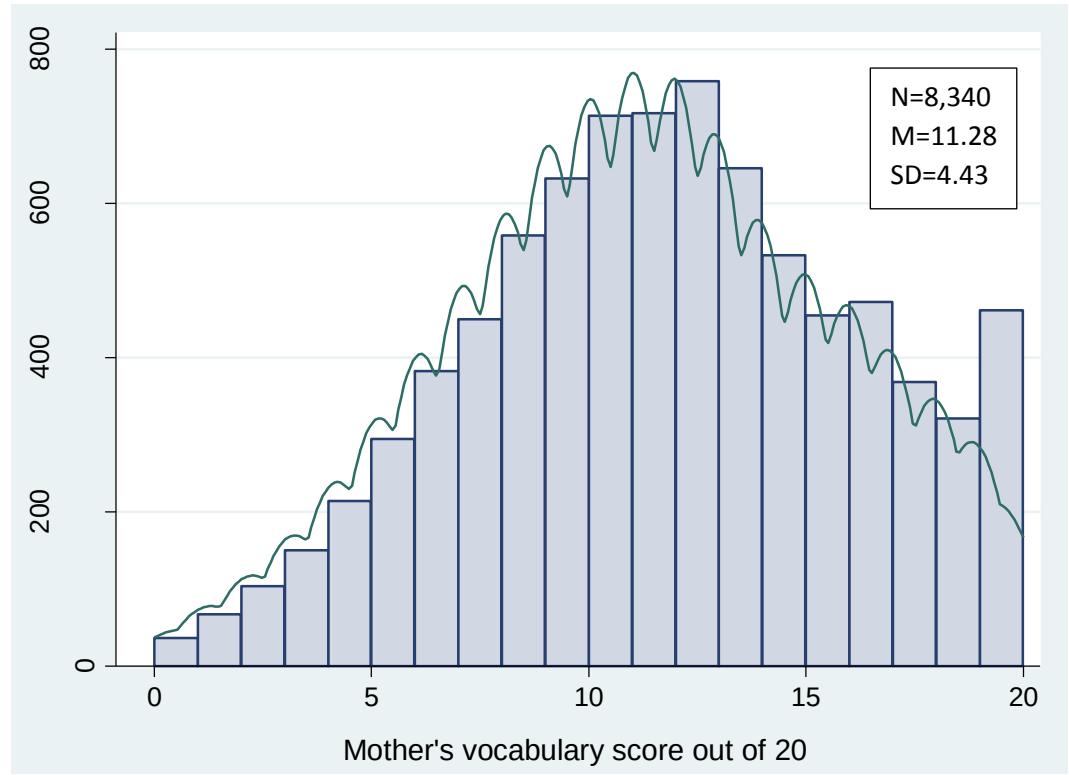

Figure 3: Distribution of partners' vocabulary scores

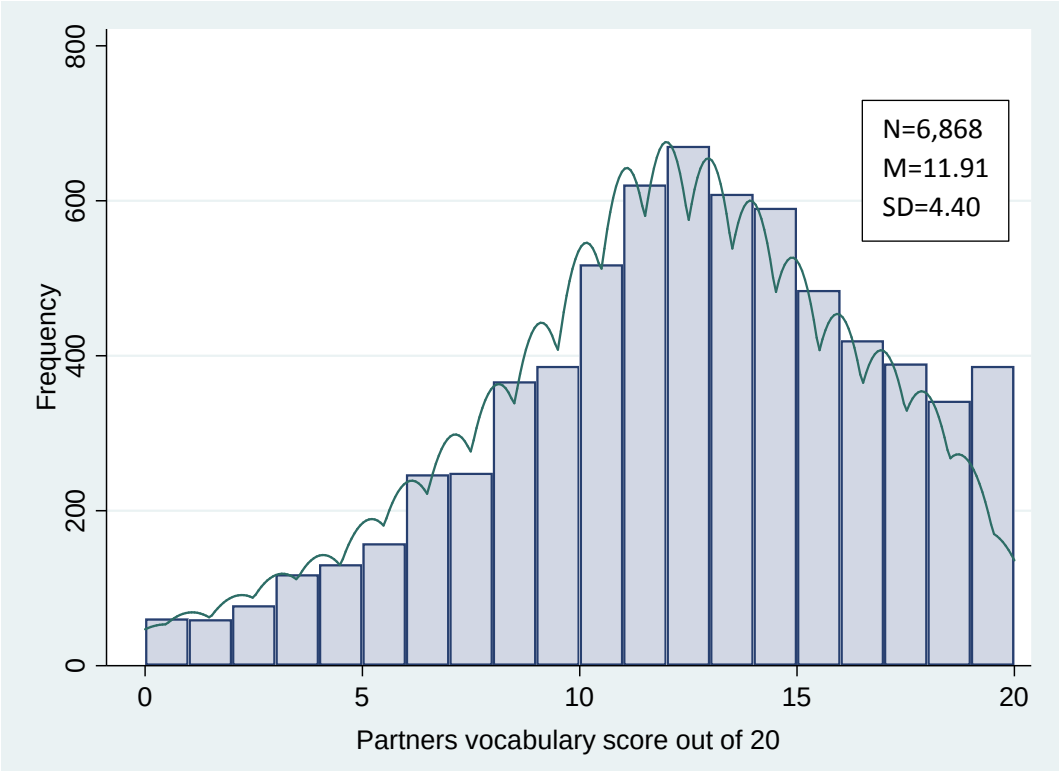

Table 1: Mean vocabulary scores for young people, mothers and partners (imputed and weighted)

	Sample characteristics %	Original N	% missing	Young person's vocabulary (mean and 95% CI)	Main vocabulary (mean and 95% CI)	Partner vocabulary (mean and 95% CI)
	n=11,714			n=11,714	n=11,714	n=8,831
Overall				6.84 (6.75, 6.93) SD=2.59	10.24 (10.03, 10.46) SD=4.42	11.17 (10.91, 11.44) SD=4.48
Age (months) quartile		11,714				
1st	27.0			6.77 (6.63, 6.90)	10.27 (10.00, 10.55)	11.05 (10.70, 11.40)
2nd	25.9			6.82 (6.67, 6.96)	10.31 (10.05, 10.57)	11.35 (11.04, 11.66)
3rd	21.9			6.95 (6.79, 7.10)	10.46 (10.13, 10.78)	11.41 (11.02, 11.79)
4th	25.2			6.85 (6.71, 6.99)	9.96 (9.69, 10.22)	10.93 (10.57, 11.28)
Gender		11,714				
Male	52.4			6.81 (6.69, 6.92)	10.22 (9.99, 10.45)	11.13 (10.85, 11.40)
Female	47.6			6.87 (6.76, 6.98)	10.27 (10.02, 10.52)	11.23 (10.92, 11.54)
Ethnicity		11,713	0.01			
White	80.5			6.91 (6.82, 7.01)	10.91 (10.71, 11.10)	11.90 (11.70, 12.11)
Mixed	4.9			6.91 (6.58, 7.24)	10.55 (10.02, 11.08)	12.12 (11.41, 12.84)
Indian	2.1			6.57 (5.96, 7.19)	7.37 (6.30, 8.44)	8.44 (7.56, 9.32)
Pakistani	3.7			6.17 (5.91, 6.43)	5.84 (5.35, 6.33)	5.76 (5.12, 6.41)
Bangladeshi	1.4			6.45 (6.14, 6.76)	4.60 (3.81, 5.40)	4.26 (3.53, 5.00)
Black Caribbean	1.6			6.32 (5.75, 6.88)	7.86 (7.04, 8.69)	10.15 (8.45, 11.86)
Black African	2.4			6.53 (6.15, 6.95)	6.85 (6.07, 7.63)	7.56 (6.10, 9.01)
Other ethnic group	3.5			6.44 (6.09, 6.78)	6.56 (5.90, 7.23)	6.74 (5.82, 7.65)
Country		11,714				
England	71.3			6.84 (6.73, 6.95)	10.43 (10.14, 10.72)	11.30 (10.93, 11.66)
Scotland	8.4			6.81 (6.61, 7.01)	10.93 (10.56, 11.30)	12.14 (11.67, 12.61)
Wales	5.0			6.59 (6.42, 6.77)	10.17 (9.84, 10.51)	10.93 (10.54, 11.32)
Northern Ireland	4.2			6.79 (6.61, 6.97)	9.85 (9.43, 10.28)	10.92 (10.42, 11.42)
London	11.1			6.96 (6.69, 7.23)	8.70 (7.88, 9.52)	9.66 (8.69, 10.64)
Household education		11,702	0.10			
No qualifications	12.7			5.81 (5.60, 6.03)	6.58 (6.14, 7.01)	7.07 (6.31, 7.84)
Vocational only	3.2			6.01 (5.63, 6.38)	7.51 (6.96, 8.06)	8.50 (7.64, 9.36)
Other academic	2.1			6.32 (5.97, 6.66)	6.44 (5.82, 7.07)	6.16 (5.22, 7.09)
GCSEd-g	9.9			6.13 (5.91, 6.36)	8.10 (7.82, 8.39)	9.27 (8.81, 9.73)
GCSEa-c	34.0			6.55 (6.44, 6.65)	9.77 (9.58, 9.95)	10.37 (10.13, 10.60)
A level or HE diploma	8.4			7.11 (6.92, 7.30)	11.64 (11.34, 11.94)	12.09 (11.74, 12.43)
Higher Diploma	10.4			7.17 (7.00, 7.34)	11.37 (11.14, 11.60)	12.04 (11.77, 12.30)
First Degree	13.9			8.13 (7.99, 8.28)	13.98 (13.71, 14.25)	14.21 (13.94, 14.49)
Higher Degree	5.3			8.63 (8.34, 8.93)	15.02 (14.62, 15.43)	15.43 (15.04, 15.82)
Household social class		11,222	4.20			
Routine	10.4			5.97 (5.68, 6.26)	7.10 (6.67, 7.52)	8.04 (7.25, 8.83)
Semi-routine	16.7			6.22 (6.05, 6.40)	8.13 (7.81, 8.46)	8.66 (8.15, 9.17)
Lower superv./tech.	10.6			6.32 (6.14, 6.49)	8.51 (8.22, 8.79)	9.30 (8.89, 9.70)
Small empl. /self-empl.	8.4			6.30 (6.09, 6.52)	8.53 (8.16, 8.90)	8.86 (8.33, 9.39)

Intermediate	11.9			6.77 (6.58, 6.96)	10.37 (10.12, 10.62)	10.81 (10.50, 11.13)
Lower manag. / prof.	25.4			7.19 (7.07, 7.30)	11.79 (11.60, 11.98)	12.29 (12.08, 12.50)
Higher manag./ prof.	16.6			8.12 (7.94, 8.29)	13.88 (13.57, 14.20)	14.61 (14.33, 14.89)

	Sample characteristics %	Original N	% missing	Young person's vocabulary (mean and 95% CI)	Main vocabulary (mean and 95% CI)	Partner vocabulary (mean and 95% CI)
	n=11,714			n=11,714	n=11,714	n=8,831
Home owner		11,698	0.14			
Yes	54.7			7.26 (7.15, 7.38)	11.64 (11.38, 11.91)	12.18 (11.89, 12.47)
No	45.3			6.32 (6.20, 6.44)	8.55 (8.34, 8.77)	9.40 (9.07, 9.74)
Income (quartile)		11,663	0.44			
1st	30.7			6.11 (5.97, 6.25)	7.96 (7.68, 8.24)	8.55 (8.08, 9.02)
2nd	26.1			6.64 (6.51, 6.77)	9.37 (9.16, 9.58)	10.02 (9.71, 10.33)
3rd	22.8			7.15 (7.02, 7.27)	11.36 (11.16, 11.55)	11.93 (11.70, 12.15)
4th	20.5			7.82 (7.66, 7.99)	13.54 (13.23, 13.84)	14.03 (13.75, 14.32)
Main age (quartile)		11,709	0.04			
1st	27.7			6.17 (6.02, 6.31)	8.32 (8.09, 8.55)	9.48 (9.12, 9.84)
2nd	26.7			6.86 (6.74, 6.98)	9.94 (9.69, 10.19)	10.83 (10.51, 11.15)
3rd	23.2			7.21 (7.06, 7.35)	11.34 (11.07, 11.60)	12.00 (11.66, 12.34)
4th	22.5			7.24 (7.09, 7.40)	11.84 (11.48, 12.19)	12.46 (12.09, 12.83)
Partner age (quartile)		8,829	0.02			
1st	22.7			6.45 (6.30, 6.61)	8.97 (8.67, 9.27)	9.62 (9.27, 9.98)
2nd	20.9			7.047 (6.84, 7.24)	10.72 (10.24, 11.20)	11.45 (11.13, 11.76)
3rd	23.6			7.05 (6.84, 7.26)	10.93 (10.46, 11.39)	11.97 (11.62, 12.33)
4th	32.7			6.82 (6.60, 7.04)	10.36 (9.75, 10.96)	12.71 (12.06, 12.37)
Home language		11,714				
English only	87.2			6.86 (6.77, 6.96)	10.77 (10.58, 10.96)	11.84 (11.63, 12.04)
Mixed or non-English	12.8			6.64 (6.45, 6.83)	6.67 (6.19, 7.16)	7.16 (6.53, 7.79)
Books in home		10,918	6.80			
<10	16.5			6.00 (5.83, 6.16)	7.15 (6.85, 7.46)	8.03 (7.48, 8.57)
11 to 25	16.0			6.34 (6.15, 6.52)	8.29 (8.00, 8.59)	9.42 (8.98, 9.85)
26 to 100	32.5			6.64 (6.53, 6.75)	10.02 (9.80, 10.25)	10.79 (10.52, 11.05)
101 to 200	17.8			7.24 (7.09, 7.38)	11.79 (11.52, 12.05)	12.50 (12.22, 12.79)
201 to 500	12.3			7.90 (7.69, 8.10)	13.60 (13.25, 13.95)	13.85 (13.53, 14.17)
> 500	4.9			8.50 (8.13, 8.86)	14.45 (13.90, 14.99)	14.95 (14.48, 15.43)
Frequency read to child (age 3)		10,706	8.61			
Not at all	3.6			5.55 (5.20, 5.91)	6.14 (5.48, 6.79)	6.55 (5.52, 7.57)
Less often	2.4			6.01 (5.60, 6.43)	7.15 (6.58, 7.71)	8.67 (7.46, 9.88)
Once or twice a month	3.3			6.16 (5.80, 6.51)	8.34 (7.74, 8.94)	9.56 (8.55, 10.57)
Once or twice a week	16.6			6.32 (6.14, 6.50)	8.64 (8.34, 8.94)	9.48 (9.09, 9.87)
Several times a week	19.6			6.69 (6.54, 6.85)	9.88 (9.63, 10.13)	10.84 (10.51, 11.17)
Every day	54.5			7.21 (7.09, 7.32)	11.38 (11.14, 11.63)	12.17 (11.92, 12.43)
YP reads (age 11)		10,833	7.52			
Never	8.2			5.85 (5.61, 6.09)	9.10 (8.73, 9.47)	9.99 (9.44, 10.55)
Less often	7.9			6.39 (6.11, 6.66)	9.81 (9.37, 10.24)	10.45 (9.95, 10.95)
At least once a month	10.9			6.53 (6.33, 6.73)	9.83 (9.49, 10.18)	10.65 (10.23, 11.08)
At least once a week	29.7			6.54 (6.42, 6.66)	9.88 (9.62, 10.14)	10.83 (10.51, 11.15)

Most days	43.3			7.38 (7.25, 7.51)	10.89 (10.60, 11.18)	11.82 (11.49, 12.16)
Plays musical instrument (age 11)		11,023	5.90			
Yes	37.1			7.33 (7.19, 7.46)	11.26 (10.95, 11.57)	12.03 (11.71, 12.34)
No	62.9			6.55 (6.45, 6.64)	9.65 (9.44, 9.85)	10.64 (10.35, 10.92)
	Sample characteristics %	**Original N**	**% missing**	**Young person's vocabulary (mean and 95% CI)**	**Main vocabulary (mean and 95% CI)**	**Partner vocabulary (mean and 95% CI)**
	n=11,714			n=11,714	n=11,714	n=8,831
Main vocabulary (quartile)		11,071	5.49			
1st	28.5			5.93 (5.79, 6.08)	5.01 (4.85, 5.17)	8.09 (7.65, 8.54)
2nd	25.7			6.53 (6.42, 6.65)	9.05 (9.01, 9.09)	10.51 (10.24, 10.77)
3rd	22.6			6.97 (6.85, 7.09)	11.97 (11.94, 12.01)	11.76 (11.54, 11.97)
4th	23.2			8.15 (7.98, 8.32)	16.29 (16.17, 16.40)	14.26 (14.00, 14.51)
Partner vocabulary* (quartile)		6,891	21.97			
1st	24.4			6.07 (5.90, 6.23)	7.63 (7.19, 8.08)	5.67 (4.58, 6.76)
2nd	24.1			6.59 (6.39, 6.78)	9.77 (9.24, 10.30)	10.48 (8.82, 11.95)
3rd	25.1			6.91 (6.70, 7.12)	10.83 (10.47, 11.18)	13.41 (11.76, 14.68)
4th	26.4			7.74 (7.32, 8.26)	12.64 (11.16, 14.11)	17.14 (16.53, 17.51)
Young person vocabulary (quartile)	10,781	7.96				
1st	17.7			3.28 (3.20, 3.35)	8.45 (8.17, 8.74)	9.64 (9.25, 10.02)
2nd	30.4			5.58 (5.56, 5.61)	9.41 (9.15, 9.67)	10.33 (10.00, 10.66)
3rd	28.9			7.50 (7.48, 7.53)	10.53 (10.27, 10.80)	11.23 (10.91, 11.55)
4th	23.0			10.39 (10.31, 10.48)	12.36 (12.01, 12.72)	13.13 (12.79, 13.46)
Verbal cognition (age 5)		10,888	7.05			
1st	21.0			5.69 (5.54, 5.84)	7.55 (7.22, 7.88)	8.10 (7.60, 8.61)
2nd	28.9			6.43 (6.31, 6.54)	9.67 (9.44, 9.89)	10.78 (10.49, 11.07)
3rd	16.5			6.86 (6.64, 7.08)	10.60 (10.20, 10.99)	11.62 (11.21, 12.03)
4th	33.6			7.90 (7.77, 8.04)	12.27 (12.00, 12.54)	12.93 (12.67, 13.19)
Picture similarities (age 5)		10,899	6.96			
1st	24.5			6.08 (5.93, 6.23)	8.87 (8.59, 9.15)	9.93 (9.53, 10.33)
2nd	18.8			6.69 (6.53, 6.85)	10.11 (9.80, 10.41)	11.01 (10.62, 11.39)
3rd	25.2			6.99 (6.83, 7.12)	10.59 (10.34, 10.85)	11.47 (11.16, 11.79)
4th	31.5			7.39 (7.24, 7.53)	11.12 (10.82, 11.41)	11.89 (11.57, 12.20)
Pattern construction (age 5)		10,449	10.80			
1st	25.1			5.98 (5.84, 6.13)	8.81 (8.54, 9.09)	9.95 (9.53, 10.36)
2nd	27.2			6.65 (6.53, 6.77)	9.85 (9.60, 10.10)	10.76 (10.44, 11.08)
3rd	22.5			7.14 (7.00, 7.28)	10.89 (10.59, 11.19)	11.57 (11.20, 11.94)
4th	25.2			7.61 (7.46, 7.77)	11.51 (11.22, 11.81)	12.29 (11.99, 12.59)
Partner response						
Partner vocab completed	53.3	11,714		7.09 (6.98, 7.20)	11.12 (10.89, 11.36)	11.72 (11.49, 11.95)
Partner vocab missing	16.8			6.61 (6.44, 6.78)	9.07 (8.67, 9.47)	9.45 (8.96, 9.93)
Not applicable(Lone parent)	29.9			6.51 (6.38, 6.64)	9.34 (9.10, 9.57)	NA

Note: Missing means imputed

Table 2 : Correlation matrix: young person's, main and partner's vocabulary and young person's early cognition

	Young person's vocabulary	Main vocabulary	Main vocabulary*	Partner vocabulary*	Verbal cognition (age 5)	Picture similarities (age 5)
Young person vocabulary						
Main vocabulary	.324 (.02) (.293, .354)					
Main vocabulary*	.321 (.02) (.286, .355)					
Partner vocabulary*	.291 (.02) (.259, .323)	.390 (.01) (.148, .683)	.544 (.02) (.509, .579)			
Verbal cognition age 5	.325 (.01) (.298, .352)	.406 (.02) (.370, .442)	.424 (.02) (.381, .467)	.414 (.02) (.369, .459)		
Picture similarities age 5	.194 (.02) (.165, .225)	.175 (.02) (.145, .205)	.189 (.02) (.152, .226)	.170 (.02) (.135, .206)	.354 (.02) (.323, .384)	
Pattern construction age 5	.242 (.02) (.213, .271)	.229 (.01) (.201, .257)	.229 (.02) (.196, .263)	.212 (.02) (.175, .248)	.372 (.02) (.343, .405)	.390 (.02) (.360, .420)
observations	11,714	11,714	8,831	8,831	11,714	11,714

*correlations for main and partner vocabulary where the partner is present in the household at MCS6 (n=8831)

Note: mean, standard error and 95% CI. Missing observations are imputed.

Table 3: Vocabulary at age 14 (standardised): Linear regression

	Model 1 Demographics		Model 2 Family resources		Model 3 Reading and music		Model 4 Parental vocab		Model 5 (final) Child cognition	
Age (months)	0.01**	(0.00)	0.01**	(0.00)	0.01**	(0.00)	0.01**	(0.00)	0.00	(0.00)
Sex (ref: boys)	0.02	(0.02)	0.01	(0.02)	-0.04	(0.02)	-0.03	(0.02)	-0.06*	(0.02)
Ethnicity (ref: White)										
Mixed	-0.04	(0.06)	-0.04	(0.06)	-0.04	(0.06)	-0.04	(0.05)	-0.01	(0.05)
Indian	-0.29*	(0.12)	-0.24*	(0.12)	-0.24*	(0.12)	-0.12	(0.12)	-0.13	(0.12)
Pakistani	-0.21**	(0.07)	-0.15*	(0.07)	-0.17*	(0.07)	-0.02	(0.07)	0.09	(0.07)
Bangladeshi	-0.09	(0.08)	-0.03	(0.08)	-0.05	(0.08)	0.17*	(0.08)	0.25**	(0.08)
Black Caribbean	-0.18+	(0.10)	-0.15	(0.10)	-0.18+	(0.10)	-0.06	(0.10)	-0.01	(0.11)
Black African	-0.18*	(0.08)	-0.13+	(0.08)	-0.14*	(0.08)	-0.00	(0.08)	0.07	(0.08)
Other Ethnic Group (inc Chinese, Other)	-0.16*	(0.07)	-0.10	(0.07)	-0.14*	(0.07)	0.01	(0.07)	0.09	(0.07)
Country (ref: England)										
Scotland	-0.08*	(0.04)	-0.07+	(0.04)	-0.06	(0.04)	-0.06+	(0.04)	-0.07+	(0.04)
Wales	-0.11**	(0.04)	-0.10**	(0.03)	-0.09**	(0.03)	-0.07*	(0.03)	-0.08*	(0.03)
Northern Ireland	0.03	(0.04)	0.04	(0.03)	0.03	(0.03)	0.06+	(0.03)	0.02	(0.04)
London	0.08+	(0.05)	0.08+	(0.04)	0.06	(0.04)	0.08+	(0.04)	0.07	(0.04)
Partner present (ref: completed vocab)										
Partner present vocab missing	-0.02	(0.03)	-0.01	(0.03)	-0.01	(0.03)	0.01	(0.03)	0.02	(0.03)
Lone parent	-0.03	(0.06)	-0.00	(0.06)	0.01	(0.06)	-0.09	(0.12)	-0.05	(0.10)
Highest Household academic qualification (ref: None)										
Vocational only	0.03	(0.09)	0.01	(0.09)	0.01	(0.08)	0.02	(0.08)	0.03	(0.08)
Other academic	0.15+	(0.09)	0.13	(0.09)	0.11	(0.09)	0.09	(0.09)	0.07	(0.08)
GCSEd-g	0.08	(0.06)	0.06	(0.06)	0.05	(0.06)	0.02	(0.06)	-0.02	(0.06)
GCSEa-c	0.18***	(0.05)	0.14**	(0.05)	0.12**	(0.05)	0.05	(0.05)	0.01	(0.05)
A level or HE diploma	0.34***	(0.06)	0.24***	(0.06)	0.22***	(0.06)	0.07	(0.06)	0.03	(0.06)
Higher Diplomas	0.32***	(0.06)	0.24***	(0.06)	0.21***	(0.06)	0.10+	(0.06)	0.04	(0.06)
First Degree	0.61***	(0.06)	0.46***	(0.06)	0.41***	(0.06)	0.19**	(0.06)	0.13*	(0.06)
Higher Degree	0.75***	(0.08)	0.56***	(0.07)	0.49***	(0.07)	0.23**	(0.07)	0.17*	(0.07)

Household NS-SEC (ref: Routine)										
Semi routine	0.05	(0.06)	0.04	(0.06)	0.04	(0.06)	0.02	(0.06)	0.02	(0.06)
Low sup and tech	0.03	(0.07)	0.00	(0.07)	0.01	(0.07)	0.01	(0.07)	0.00	(0.07)
Small emp and s-emp	-0.02	(0.07)	-0.03	(0.07)	-0.03	(0.07)	-0.03	(0.07)	-0.05	(0.07)
Intermediate	0.10	(0.07)	0.06	(0.07)	0.07	(0.07)	0.02	(0.07)	0.01	(0.07)
Lo manag/prof	0.11	(0.07)	0.06	(0.07)	0.06	(0.07)	-0.01	(0.07)	-0.04	(0.07)
Hi manag/prof	0.24**	(0.07)	0.16*	(0.07)	0.15*	(0.07)	0.04	(0.07)	0.01	(0.07)
Home owner (ref: do not own home)	0.01	(0.03)	0.01	(0.03)	0.02	(0.03)	0.01	(0.03)	-0.00	(0.03)
Household income (log)	0.04+	(0.02)	0.04+	(0.02)	0.04+	(0.02)	0.02	(0.02)	0.01	(0.02)
Mother's age	0.01*	(0.00)	0.00+	(0.00)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
Partner's age	0.00	(0.00)	-0.00	(0.00)	-0.00	(0.00)	-0.00	(0.00)	-0.00	(0.00)
Number of older siblings	-0.06***	(0.01)	-0.06***	(0.01)	-0.05***	(0.01)	-0.05***	(0.01)	-0.03*	(0.01)
Number of younger siblings	-0.02	(0.01)	-0.02	(0.01)	-0.02	(0.01)	-0.02	(0.01)	-0.01	(0.01)
Speak other language at home (ref: English only)	0.12*	(0.06)	0.14**	(0.05)	0.12*	(0.05)	0.22***	(0.06)	0.27***	(0.05)
Number of books at home (ref: <11)										
11-25			0.07	(0.05)	0.07	(0.05)	0.03	(0.05)	0.02	(0.04)
26-100			0.08*	(0.04)	0.07+	(0.04)	-0.00	(0.04)	-0.02	(0.04)
101-200			0.21***	(0.04)	0.18***	(0.04)	0.05	(0.04)	0.02	(0.04)
201-500			0.36***	(0.05)	0.31***	(0.05)	0.12*	(0.05)	0.09+	(0.05)
More than 500			0.51***	(0.07)	0.46***	(0.07)	0.25***	(0.07)	0.22***	(0.06)
Frequency reads to child (age 3) (ref: Never)										
Less often			0.15	(0.10)	0.15	(0.10)	0.14	(0.10)	0.09	(0.10)
Once or twice a month			0.14	(0.10)	0.13	(0.10)	0.09	(0.10)	0.04	(0.09)
Once or twice a week			0.17*	(0.08)	0.16*	(0.08)	0.13+	(0.07)	0.07	(0.07)
Several times a week			0.20*	(0.08)	0.18*	(0.08)	0.14+	(0.08)	0.09	(0.07)
Every day			0.26***	(0.08)	0.22**	(0.08)	0.17*	(0.08)	0.09	(0.07)
Reads for pleasure (age 11) (ref: Never)										
Less often					0.15*	(0.06)	0.13*	(0.06)	0.09	(0.06)
At least once a month					0.18**	(0.06)	0.18**	(0.06)	0.12*	(0.06)
At least once a week					0.16***	(0.05)	0.15**	(0.05)	0.11*	(0.05)
Most days					0.37***	(0.05)	0.34***	(0.05)	0.28***	(0.05)
Plays musical instrument (ref: does not play)					0.09***	(0.02)	0.09***	(0.02)	0.06**	(0.02)
Main vocab score (zscore)							0.17***	(0.02)	0.14***	(0.02)

Partner vocab score (zscore)							0.13***	(0.02)	0.10***	(0.02)
Verbal cognition zscore (age 5)									0.18***	(0.01)
Picture similarities zscore (age 5)									0.03**	(0.01)
Pattern construction zscore (age 5)									0.09***	(0.01)
Constant	-2.41***	(0.53)	-2.67***	(0.53)	-2.86***	(0.52)	-2.30***	(0.51)	-0.84+	(0.50)
Observations	11,714		11,714		11,714		11,714		11,714	

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05, + p<0.10
Multiple imputation was applied to all missing data, including absent partners in single family households at MCS6 and controlled for by including a Partner present variable in the model