Longitudinal and Life Course Studies • vol 13 • no 1 • 87–120

Online ISSN 1757-9597 • https://doi.org/10.1332/175795921X16233448663756

RESEARCH ARTICLE

Development of socio-economic gaps in children's language skills in Germany

Thorsten Schneider, thorsten.schneider@uni-leipzig.de
University of Leipzig, Germany

Tobias Linberg, tobias.linberg@las.bayern.de
Bavarian State Institute for School, Germany

Gaps in language skills by socio-economic status (SES) are already evident before school entry, and these gaps may change over time. After discussing mechanisms of cumulative advantages ('Matthew effects') and compensatory effects as well as the relevance of cultural capital and child-related activities in families, this paper tests mechanisms behind changing SES gaps in language skills from age five to nine in Germany. Analysing data from the German National Educational Panel Study with growth curve models, we find widening SES gaps in children's vocabulary. Children of mothers with low educational attainment show a far below-average increase in skills. The findings are in line with cumulative advantage by status, although initial skills predict their growth over time as well. There are no signs of any type of compensatory effects. Reading aloud to children appears to substantially impact and mediate SES differences in vocabulary progress.

Key words Matthew effects • compensatory advantages • receptive vocabulary • social inequality • growth curve models

Key messages

- Based on longitudinal data, vocabulary development from age five to nine in Germany is researched.
- At age five, vocabulary skills of children with low- and high-educated mothers differ by one year.
- This gap doubles in the following four years, hinting at cumulative advantages.

To cite this article: Schneider, T. and Linberg, T. (2022) Development of socio-economic gaps in children's language skills in Germany, *Longitudinal and Life Course Studies*, vol 13, no 1, 87–120, DOI: 10.1332/175795921X16233448663756

Introduction

Language skills are considered a key factor in educational success: learning in school is dependent on language skills that are closely linked to competencies and progress in reading (Sénéchal et al, 2008) and also to memory performance (Weinert, 2010) and metacognitive competencies (Ebert, 2011). Receptive vocabulary, which comprises the words we know, is a central marker of language development and evolves rapidly in the preschool and elementary school years (Taylor et al, 2013: 1). This sharp increase in language skills at preschool age and during the first few years of elementary school calls for a careful assessment of the underlying mechanisms that stimulate this growth, especially in presence of social inequalities.

Bioecological and transactional models of child development describe how individual development is intertwined with contextual factors when individual variation is explained (Shonkoff and Phillips, 2000; Bronfenbrenner and Morris, 2006; Sameroff, 2010). There are large individual differences in language skills during early childhood, and the accumulated results of educational and psychological research show that this individual variability is driven in part by contextual factors (Vasilyeva and Waterfall, 2011). From the perspective of sociology, whose core focus is on the causes and consequences of social stratification, the crucial contextual factors determining children's development and life chances are parents' education, social class and status (Bukodi and Goldthorpe, 2013), which are closely related to families' resources and activities. Hereafter, we use socio-economic status (SES) as a generic term to capture such differences in families' position in a society as well as in resources available to them.

There is abundant evidence in sociology, but also in other disciplines, that substantial differences by SES exist in language skills even before children enter elementary school (Lee and Burkam, 2002; Sammons et al, 2007; Weinert and Ebert, 2013; Fernald et al, 2013; Garcia, 2015; Linberg and Wenz, 2017; Linberg et al, 2019; McMullin et al, 2020) and that impacts of early educational investments and decisions on skill development and educational attainment are long-lasting (Bynner, 2004; Cunha et al, 2006; Heckman and Masterov, 2007).

In short, it can be concluded that language skills are important for educational success, develop rapidly during childhood, show a wide range in individual variability, are influenced by contextual factors and differ by SES already at preschool age. However, longitudinal studies on the development of these inequalities are less common, especially when underlying mechanisms are investigated this early. As early (dis)advantages can have lasting consequences (Alexander et al, 2007), the first aim of this paper is to describe if and how SES gaps in early language skills change throughout early childhood in Germany. The second aim is to gain a better understanding of whether and how activities within the family shape the development of SES gaps.

We focus on Germany for several reasons: in cross-national comparisons, social disparities in reading skills are highly pronounced in Germany at age 15 (OECD, 2010: 16), and these disparities are largely driven by the large proportion of students from educationally disadvantaged families who remain at the lowest competence levels up to the end of their compulsory schooling (for example, Klieme et al, 2010). Recent research shows that children from families with low education are already heavily over-represented among students with lower scores on language tests at the age of five (Linberg and Wenz, 2017). The German National Educational Panel

Study (NEPS) was the first nationwide longitudinal study on educational processes and competence development in Germany and provides a representative sample of children who have been tested three times in the same domain to address the issues under investigation here.

The paper is organised as follows. In the next section, we take a closer look at cumulative advantages ('Matthew effects') following some distinctions proposed by DiPrete and Eirich (2006). We then discuss issues of compensatory effects and compensatory advantages by status, an idea put forward in sociology more recently by Bernardi (2014). To shed more light on the link between parents' SES and children's skill development, we consider sociological approaches to cultural capital and concerted cultivation (Bourdieu, 1984; 1986; Lareau, 2011). After explaining how these mechanisms can be identified, we apply these ideas to the research on early inequalities by testing the different mechanisms with growth curve models using data from the NEPS (Blossfeld and Roßbach, 2019). Aims, and research questions are presented next, then data and methods, and results and conclusions.

Theoretical approaches and empirical findings on changing SES gaps

Cumulative advantages

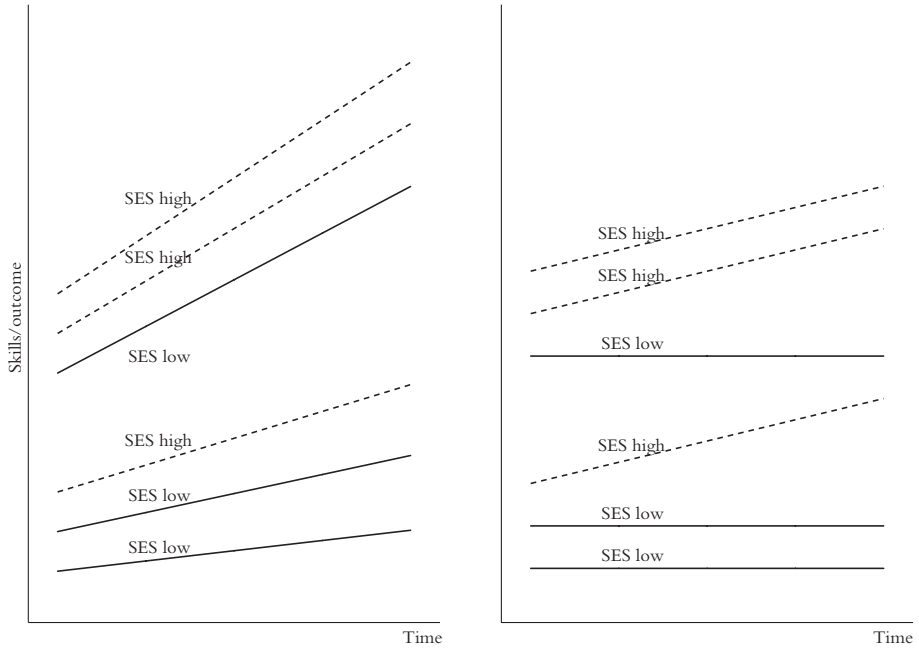
The idea that advantages such as skills accumulate over time is widespread in the research on social stratification and educational inequalities across disciplines (building on early contributions by Merton, 1968; 1988; Walberg and Tsai, 1983; Shaywitz et al, 1995). Also known as the Matthew effect,[1] the concept of cumulative advantage describes the process by which the advantages of privileged individuals or groups increase while the less privileged lag further behind, resulting in a widening gap over time (see Figure 1).

Cumulative advantage in its strict form

DiPrete and Eirich (2006) spelled out different types of cumulative advantage[2] (CA) in sociology. In the strict form of cumulative advantage, the level of resources already available influences the level of resources that are subsequently attained. That is, individuals or groups might make more progress than others because their higher starting capital leads to more revenues (per unit) or attracts disproportionately high attention, leading to faster promotion or more resources and, in turn, to more capital. Here, capital is a generic term and comprises different types of resources such as wealth, recognition, skills and abilities. When the advantage itself pushes the progress, there is a path-dependent process. DiPrete and Eirich (2006: 276–7) use the term *strict form of cumulative advantage* for such cases. Here, 'CA becomes part of an explanation for growing inequality' (DiPrete and Eirich, 2006: 272) rather than a description of distributions over time or the life course. The *strict form of cumulative advantage* equals to *self-productivity* in the economic model of skill formation (Cunha et al, 2006: 702–3).

Transferring these thoughts to early skill development, children who start off with higher skills and endowments are expected to show a steeper development because high skill endowments may contribute both to more effective learning and to processes of reinforcement in certain (selected and selective) contexts. Research

Figure 1: Two types of fan-spread effects: (left) strict form of cumulative advantage vs (right) increasing SES gaps in skills due to 'exposure'



Notes: Dashed line: children from high-SES families; solid line: children from low-SES families.

Source: authors' illustration.

indicates increasing differences in specific types of capital and in specific domains. With respect to reading, Stanovich (1986) reported that children who have a head start in reading skills also show more progress over time, because higher reading proficiency increases phonological awareness, vocabulary and word recognition, which in turn should foster reading progress. Thus, reciprocal relationships and positive feedback can be expected (Stanovich, 1986: 364).[3] More recent findings from the Netherlands on the reciprocal relationship between vocabulary and reading skills in elementary school support this explanation (Verhoeven et al, 2011). Duzy et al (2014) provide further support for Matthew effects in the relationship between word decoding and text comprehension in German elementary school students. The idea of cumulative advantage in its strict form is shown in the left panel of Figure 1. The increase in skills only depends on 'initial' (or previous) skill levels. However, two SES groups are differentiated in this figure, and initial skills should be higher for high-SES children on average. In Figure 1, one can see an increase in SES gaps over time, but this increase is not due to SES, but due to differential rates of progress by initial skills.

Broader concepts of cumulative advantage

DiPrete and Eirich (2006) also discuss broader concepts of cumulative advantage used in theoretical and empirical research, focusing on accumulated (dis)advantages, repeated discrimination by status over the life course, or 'exposures to a treatment over some (possibly long) duration' (DiPrete and Eirich, 2006: 273). Social reproduction

theory and the theory of concerted cultivation are among these concepts. Bourdieu (1984; 1986), a major proponent of social reproduction theory, argued that cultural capital is acquired first and foremost in the family, as parents with higher cultural capital and greater financial resources create more stimulating contexts for their children's development. Research across different disciplines in social sciences has repeatedly shown that provision of language stimulation varies substantially by SES (for example, Hart and Risley, 1995; Pellegrini, 2002; Hoff, 2006b). Lareau (2011) theorised and showed that higher-SES parents employ a 'concerted cultivation' style of parenting, leading to an increasing sense of belonging and entitlement in children within the school context and growing differences in children's everyday lives by social background. In an endeavour to operationalise Lareau's ethnographic concept, Cheadle (2008) used several indicators to capture 'concerted cultivation' in a single factor.[4] This factor is positively associated with an increase in reading skills during kindergarten and first grade, but not later on. He concluded: 'While concerted cultivation plays a role in explaining socio-economic advantage, it is only a modest part of the story' (Cheadle, 2008: 23). In addition, Pensiero (2011) showed that the effects of concerted cultivation on children's reading skills can be traced back to stimulating processes and reading activities and not to participation in organised leisure activities.

There is abundant, but not absolutely consistent evidence of the effects that proximal processes in children's immediate environments have on language skills. Such proximal processes can be found in stimulating activities that imply language input (Hoff, 2006a; Vasilyeva and Waterfall, 2011; Schmerse et al, 2018) and especially in (shared) book reading activities (Scarborough and Dobrich, 1994; Mol et al, 2008; Klein and Kogan, 2013). According to Taylor et al (2013), however, these processes did not show a longitudinal effect on the growth of language skills. A very recent study on vocabulary development between the ages of three and five in Ireland underscores that not all findings are consistent. Home learning activities and test results clearly correlate positively at age three. In a longitudinal analysis on gains in expressive vocabulary, however, only children from families in the lowest quintile of activities lagged somewhat behind. The other coefficients are far from being statistically significant although the sample consists of more than 8,000 children (McMullin et al, 2020: 618).

The major difference between the two aforementioned approaches to cumulative advantage is that according to the stricter concept, the amount of capital (resources, skills) under investigation is itself thought to cause an increase in disparities, whereas the broader concept refers to other (exogenous) factors such as stratified resources and activities that drive the increase in differences. We refer to this as the 'exposure model' and visualise the idea in the right panel of Figure 1. Here, progress in skills differs by SES, independent of the initial (or previous) skill level.

Compensatory effects and advantages

In contrast to cumulative advantages, which lead to increasing gaps, compensatory effects occur when differences in skills or other resources decrease within a population. One can distinguish two types of compensatory effects: those in which the spread of the overall distribution shrinks, and those in which the spread shrinks only within specific subpopulations.

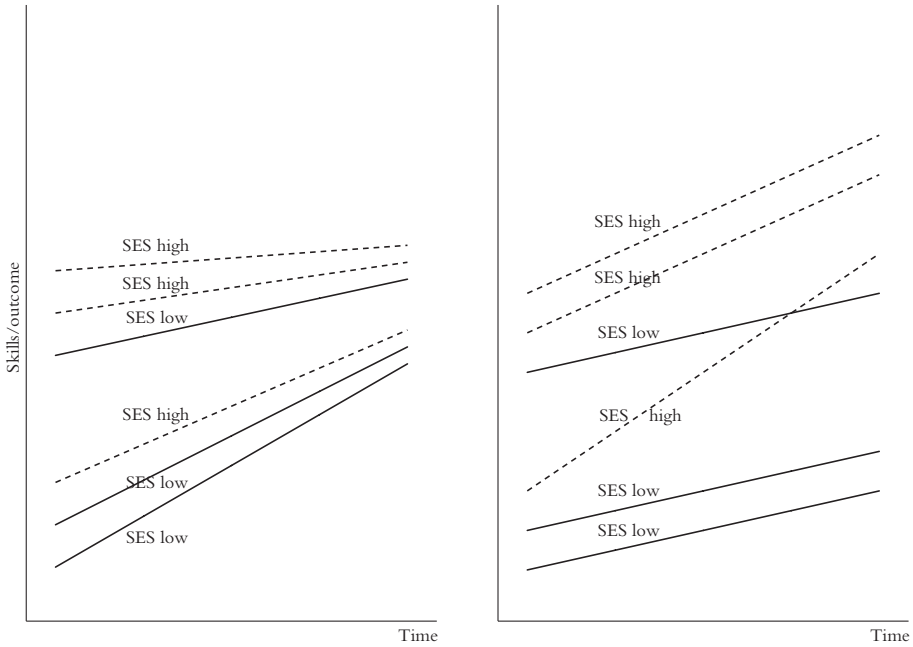
The former type can be referred to as *pure compensatory effects*. In line with this idea, almost all children can be assumed to acquire some basic skills such as knowledge of letters or the Roman alphabet (Pfost et al, 2014: 209). Those who progress faster in the beginning might master these basic skills earlier, but those with lower initial skills or 'late starters' might catch up at some point, leading to a shrinking or even completely vanishing spread within the whole population over time (Pfost et al, 2014). In addition, early child education and care and, later on, school might reduce the spread of the overall distribution by targeting and promoting children who lag behind and offering them a more stimulating, higher-quality 'instructional regime' (Raudenbush and Eschmann, 2015) compared to their family environment.

Compensatory effects might also vary, however, by SES. This latter type of compensatory effect, in which the spread shrinks within specific subpopulations, can be referred to as *compensatory advantage*. This is the case if not all children with low initial skills catch up, but only or primarily those from higher-SES families. According to Feinstein (2003: 73), 'the children of educated or wealthy parents who scored poorly in the early tests had a tendency to catch up, whereas children of worse-off parents who scored poorly were extremely unlikely to catch up and are shown to be an at-risk group'. This study has been subjected to critical analysis (Jerrim and Vignoles, 2013) and intense debate (Feinstein et al, 2015), as the reported pattern may also result from a statistical artefact called 'regression to the mean'. However, using more measurement points and more sophisticated statistical methods, Bradbury et al (2015: 121–4) found support for Feinstein's previous findings in a study on children from kindergarten to eighth grade in the United States: children with low reading skills from high-SES families caught up over time, whereas children with high reading skills from low-SES families lost ground. This effect can be traced by looking at group-specific changes in the skill distribution over time: if especially high-SES children with initially low skills catch up over time, the skill distribution of this high-SES group becomes comparatively more homogeneous over time.

Further support for compensatory advantages comes from Bernardi and Grätz's (2015) research on the effect of birth month on later school achievement in England. Here, the negative effect of being one of the youngest children in a school entry cohort on later achievement measured at age 11, 14 and 16 is especially strong for students from low-SES families, but not for students from high-SES families. Correspondingly, Bernardi (2014) describes socio-economic variations in path dependency on prior negative outcomes as a central insight into the concept of compensatory advantage: 'Put another way, chances of success at $t + 1$ are more dependent on the previous outcome at time t for the working class than for the upper class' (Bernardi, 2014: 75). However, one should bear in mind that mechanisms behind compensatory advantages might be the same as those already discussed in the context of cultural capital, concerted cultivation and direct language stimulation.

Again, the major difference between the two approaches is that the idea of *pure compensatory effects* refers to the amount of capital (resources, skills) under investigation itself as reason for a decrease in disparities, whereas *compensatory advantage* refers to some other (exogenous) factors that drive the decrease in differences over time in a specific subpopulation. Both approaches are shown in a stylised way in Figure 2. While in the case of pure compensatory effects, skill differences narrow and SES gaps should therefore shrink (left panel), in the case of compensatory advantage, differences

Figure 2: 'Pure' compensatory effects (left) vs compensatory advantage by SES (right)



Notes: Dashed line: children from high-SES families; solid line: children from low-SES families.

Source: authors' illustration.

within at least one group – here, children from high-SES families – shrink, but not differences in the other group. At the aggregate level, SES gaps increase over time (right panel).

Overview of findings from longitudinal studies on inter-individual differences in skills and the development of SES gaps

Research on cumulative advantage in its pure form – which means path dependency, that is, a positive correlation between 'initial skills' and progress and therefore increasing population differences in skills over time – shows some age- or educational-stage-specificity. Baumert et al (2012: 1350) summarises previous results reported in the literature: 'Findings consistent with the Matthew effect have been found only for early phases of competency acquisition. During the school years, in contrast, students' reading development has been observed to show compensation processes'. With regard to vocabulary growth, Rowe et al (2012) report substantial differences between children with respect to initial size of vocabulary and growth rates. Similarly, the results of Rice and Hoffman (2015) document strong growth and variation within individuals, but also high stability across ages (from 2½ to 21 years) when looking at the rank ordering of individuals.

Although findings from longitudinal studies focusing on changes in SES gaps differ to some extent across samples, countries, skill indicators and age groups, the majority of research findings point to rather increasing than decreasing SES gaps.

Schmerse et al (2018) report increasing SES gaps in vocabulary and grammatical skills from the ages of 34 to 48 months in Germany. Taylor et al (2013) report stable SES differences in vocabulary development from ages four to six and eight using data from the Longitudinal Study of Australian Children. The effect is especially apparent at four years of age but relatively negligible for predicting further change. Based on US data from the Early Childhood Longitudinal Study, Cheadle (2008) reports increasing SES differences in reading skills from kindergarten to grade 1. Rice and Hoffman (2015) studied vocabulary development over a long time period (from age 2½ to 21) and report a small but noteworthy cumulative effect over time for maternal education in a US sample. Bradbury et al (2015) report slightly increasing SES differences in language and reading skills from age 5 to 11 in Australia and the United States as well as in Canada from age 5 to 7, but slightly decreasing gaps from age 7 to 11 in the United Kingdom. A study from grades 3 to 6 using data collected in the German capital reveals some differences with respect to the type of language skills. While the gaps in text comprehension do not widen with respect to the number of books in the household or parents' scores on the International Socio-Economic Index (ISEI), gaps in receptive vocabulary increase with respect to the first indicator, books. Differences in receptive vocabulary, however, are stable with respect to ISEI (Kigel et al, 2015). When looking at between-country differences in results, it remains unclear whether these can be explained by the different SES indicators or whether this suggests that SES does have different implications across countries.

Research questions

The aim of this study, as already stated, is to investigate how SES gaps in early language skills change throughout early childhood in Germany and to gain a better understanding of the development of these gaps. We add to the existing research by deriving analytical models that enable us to disentangle the major theoretically derived mechanisms presented earlier and by using data from a recent nationwide cohort of young children in Germany, where social disparities in developmental outcomes and school achievement have been shown to be substantial (see: Klieme et al, 2010; Linberg et al, 2019). The analyses are structured around the following research questions (RQ):

Are cumulative advantages present in strict form (RQ 1.1) or by exposure (RQ 1.2) based on vocabulary development from age five to nine?

- If cumulative advantages are present in strict form (RQ 1.1), the resulting picture would resemble a fan spread (left panel of Figure 1), as children with higher vocabulary skills should make more progress over time than children with low skills. In this case, the overall distribution of skills would increase. As a side effect, the raw SES gaps would have to increase, too, as higher-SES children should show initially higher skills on average.
- With regard to the idea of cumulative advantages by exposure (RQ 1.2), favourable family conditions should have ongoing positive exposure-driven impacts. In the case of this 'exposure model', children from high-SES families would show a steeper increase in their skills over time than their lower-SES counterparts (right panel of Figure 1). However, the initial or previous skill level would not be the main driving force.

Do we find evidence of pure compensatory effects (RQ 2.1) or compensatory advantages (RQ 2.2)?

- In the case of pure compensatory effects (RQ 2.1), the overall variation in skills would decrease over time (left panel of Figure 2), as the initial skill level would correlate negatively with the rate of progress. As a side effect, the SES gaps in skill distribution would shrink.
- In the case of compensatory advantages by SES (RQ 2.2), high-SES children with initially low skills would show a steeper increase in skills compared to their lower-SES counterparts (right panel of Figure 2), as higher-educated parents can be expected to invest more resources to compensate for low initial skills. The skill distribution would only or especially shrink in the group of high-SES children.

Finally, we focus on the role of family activities and ask what kind of family activities impact vocabulary progress above and beyond SES (RQ 3.1) and are these activities associated with increasing SES gaps over time, as social reproduction theory or the theory on concerted cultivation suggest (RQ 3.2)?

Data

The German national educational panel study

We use data from a cohort of the NEPS, which started in 2011 with 2,949 approximately five-year-old children attending a German kindergarten, which we refer to here as preschool (271 preschools).[5] The nationwide preschool sample 'was established using a two-stage indirect sampling approach' (Steinhauer et al, 2016: 4) by drawing, in the first stage, a probability sample of elementary schools, which were asked to list the preschools that feed into them. In the second stage, a further probability sample was drawn from the list of preschools. In both cases, sampling probabilities varied by the size of the institution. Preschools were then asked to list all children they expected to enrol two years later, and the parents of these children were asked to participate and to allow their children to participate in testing (Steinhauer et al, 2016).

Besides gathering information from parents by telephone interview (CATI) and from teachers by questionnaire, each child was tested individually in different domains. A subsample of 576 children was also group-tested during elementary school (Steinhauer et al, 2016: 5). As we aim to trace developments in language skills, we focus on this longitudinal subsample. As vocabulary tests were run in wave 1 ('preschool, approx. age 5'), 3 ('grade 1', approx. age 7), and 5 ('grade 3', approx. age 9), we restrict the longitudinal subsample to those children who have a valid test result in wave 1 plus at least one further valid test result in wave 3 or 5. This leads to a longitudinal sample of 556 children, with 528 children in wave 3 and 473 in wave 5. On average, there are 2.8 valid test results per child in our sample. We opt for an unbalanced panel design, as the share of partial unit non-response is low, and a balanced panel design is less efficient (Curran et al, 2010: 125; see also Raudenbush and Bryk, 2002: 199).

Compared to the first wave, the longitudinal subsample is selective with respect to some core socio-demographics. In particular, the share of immigrant families (defined

later) is lower in the longitudinal subsample (37% vs 21%).[6] The share of mothers holding a tertiary degree is slightly higher in the longitudinal subsample (20% vs 24%) as is the share of girls (49% vs 51%). In the longitudinal sample, there is little variation over time (share of immigrants before and after imputation at wave 1: 22%, wave 3: 21%, wave 5: 22%), mothers holding a tertiary degree (24%, 24%, 25% before, 23%, 24%, 25% after imputation), and the share of girls (51%, 51%, 50%, no missing information).

Dependent variable: receptive vocabulary skills

Children's vocabulary skills were measured with an adapted version of the Peabody Picture Vocabulary Test (PPVT; Dunn and Dunn, 2004). In this test, children listened to recorded words and were asked to select the corresponding picture from four choices. The test comprised 77 items and was conducted individually in children's preschools in wave 1. It comprised 66 items in wave 3 and 71 items in wave 5 and was conducted in both waves in groups in the children's elementary schools. In wave 1, the test instructor wrote down the children's answers, and in later waves, the children filled out the test forms themselves. The selection of items was based on item difficulties and on data from previous studies (Berendes et al, 2013).

As sets of common items were used throughout all three waves,[7] we applied item response theory (IRT) to determine children's latent vocabulary skills and to scale the test data for longitudinal use with Stata 16.1.[8] A two-parameter model allowed item difficulty and item discrimination to vary. In this framework, the common items, also called anchor items, were used to put test results from different waves on a common metric. Tests on differential item functioning revealed that some anchor items varied in their difficulty conditional on children's skills in different waves. However, skill parameters estimated by separate, wave-specific models correlated highly ($> .99$) with those estimated with a common model. Therefore, differential item functioning could be ignored. Due to iteration problems that accompanied the high number of resulting parameters (= 246), items that discriminated similarly were grouped together in the final model, allowing them to have their own difficulty but to share a common discrimination value (grouping was done in 0.4 steps from 0 to 2.4 plus a further group for items with discrimination scores above 2.4). The latent parameter on vocabulary skills ranges from −2.82 to 2.82, with a mean of 0 and a standard deviation of almost 1 ($M = -0.00$, $SD = 0.96$).

As an indicator for time, we used age in years and centred all age information to the median of wave 1, resulting in the following means and standard deviations for waves 1, 3 and 5: $\overline{Age_1} = 0.01$ years, $SD_1 = 0.32$; $\overline{Age_3} = 2.00$ years, $SD_3 = 0.31$; $\overline{Age_5} = 3.74$ years, $SD_5 = 0.32$.

Indicator of socio-economic status

With regard to SES indicators, it has been recommended to use single indicators rather than composite indicators in research on early inequalities (Duncan and Magnuson, 2003). Among the individual SES indicators, parental education shows stronger relationships with children's test scores (Cheadle, 2008) and later educational attainment (Bukodi and Goldthorpe, 2013) than occupational prestige or income. Furthermore, 'parents with more education tend to make longer utterance to their

children and use more complex language (Hoff, 2003), hold higher academic expectations, and provide more cognitive stimulation (Davis-Kean, 2005; Magnuson and Duncan, 2006)' (Quinn, 2015: 122). Looking at mothers' school and university education, we used a shortened version of the CASMIN scheme (Braun and Müller, 1997; Kerckhoff et al, 2002) and differentiated low (CASMIN 1a, b, c), intermediate (CASMIN 2a, b), upper secondary (CASMIN 2c_gen, d) and tertiary education (CASMIN 3a, b). The first group covers mothers with no or the lowest school degree, the second mothers with an intermediate school degree, the third mothers with a university entrance qualification but no tertiary education and the fourth mothers with a tertiary degree.

Indicators of cultural capital and activities at home

To test the importance of cultural capital and potentially stimulating activities at home, we considered *household possessions*, consisting of (1) the number of books (six categories from '0 to 10 books' to 'more than 500 books'), and yes/no answers on possessing (2) classic literature, (3) dictionaries, (4) poetry, (5) a library card or (6) artwork measured at wave 1 (all items standardised, Cronbach's alpha = 0.61, mean = 0, SD = 1); parents'[9] *reading time* on a work day and on a day off, measured in hours and minutes in waves 1 and 3 (all four items standardised, Cronbach's alpha = 0.76, mean = 0, SD = 1); frequency of *literacy-related activities* at home with child, namely (1) reading aloud to child, (2) using picture books, word puzzles and similar, (3) learning the alphabet, (4) memorising poems, rhymes or songs, measured with eight categories ranging from 'never' to 'several times a day' in wave 1 (all items standardised, Cronbach's alpha = 0.56, mean = 0, SD = 1); and frequency of *reading aloud* to the child as a single item measured in waves 1, 2 and 3 (all items standardised, Cronbach's alpha = 0.65, mean = 0, SD = 1). In the case of parents' own reading time and reading aloud to children, we also used information gathered in later waves to reduce measurement error. Of course, the 'appropriateness' of activities – such as learning the alphabet – depends on the child's age and skills. However, we assume that parents will adapt their literacy-related activities over time in such a way that those who provided stimulating activities at earlier ages also are more likely to provide stimulating environments later on.

Control variables

In the analyses, we controlled for the child's *gender* (girl = 1, boy = 0), non-German *language at home* (yes = 1, no = 0; provided by teachers), immigration status (yes = 1 if at least one grandparent born abroad, no = 0; Kristen et al, 2016), and the logarithm of the *number of children* under age 14 in the household as an indicator of potential dilution of resources (Downey, 2001).

Methods

Missing values and imputation

To avoid a loss of cases due to item non-response, we used information on the mother's education and the number of children in the household surveyed in wave 2

if data were missing in wave 1. If parents did not report the child's gender or birth date, we used information provided by the preschool. Multiple imputation with chained equations was applied to remaining cases of item non-response. In addition to the variables used in the analyses, the imputation model includes a set of auxiliary variables.[10] Imputations on vocabulary skills, our dependent variable, are discarded and set to missing before performing the growth curve models. In order to take the variance between estimations of different imputed data sets into account, we calculate standard errors according to Rubin's (2004) rules. Table 1 provides key information on all variables used in the main analysis before and after imputation, including the proportion of missing data.

Analytical procedure: growth curve model

The growth curve model allows us to model differences in skill progress between children and to separate between initial skills and progress in skills. We are interested in both issues, as we want to know if progress in skills varies between children and if this variation depends on SES as well as if there are real path-dependencies. Frequently used value-added models, regressing test results on previous test results and some other covariates, are incapable of separating these two effects (Raudenbush and Bryk, 2002: 166).

Data are nested as they contain information on vocabulary skills (Y_{ti}) from students ($i = 1, \ldots n$) tested in different waves ($t = 1, 3, 5$). Because children differ in their age at measurement, and vocabulary skills are age-dependent (correlation between age and skills in each wave: $r_{t=1} = 0.30$; $r_{t=3} = 0.18$; $r_{t=5} = 0.15$, before imputation), we use age at the measurement point as the key time-varying covariate on the lowest level (level 1; see eq. 1).

$$\text{Level1: } Y_{ti} = \pi_{0i} + \pi_{1i} Age_{ti} + e_{ti} \tag{1}$$

$$\text{Level2 (random intercept): } \pi_{0i} = \beta_{00} + r_{0i} \tag{2}$$

$$\text{Level2 (random slope): } \pi_{1i} = \beta_{10} + r_{1i} \tag{3}$$

For the growth curve model, at least two further equations are needed at the individual level (level 2): one for the random intercept and one for the random slope, which is the age coefficient in this case. In the simplest form, the two can differ between individuals: the intercept, representing vocabulary skills at 'baseline age' ($Age_{ti} = 0$), and the degree of progress (see eq. 2, 3). Initial differences between individuals and individual-specific differences in progress are captured by the random error terms r_{0i} and r_{1i}. Differences between the observed and predicted vocabulary skills at Age_{ti} in individuals are captured by the error term e_{ti} in equation 1.

We specified models that allow the errors r_{0i} and r_{1i} to be correlated and devote attention to this correlation. Assuming that vocabulary skills grow over time ($\beta_{10} > 0$), a positive correlation between the two error terms indicates increasing achievement inequality between children. This becomes evident, looking at equations 2 and 3. Children with initially high skills, that is, π_{0i} above β_{00}, have positive values for r_{0i}, those with lower skills, negative values for r_{0i}. Children with above-average progress, $\pi_{1i} > \beta_{10}$, show positive values for r_{1i}, those with below-average progress, negative values for r_{1i}. If

Table 1: Means and standard deviations (SD) in brackets for metric covariates and frequencies for categorical covariates, plus proportion of missing data

Variables	Mean (SD)		
	before	after	Missing
	Imputation		
Vocabulary skills			
at wave 1 (n = 556)	−0.85	−0.85	n.a.
	(0.56)	(0.56)	
at wave 3 (n = 528)	0.10	0.10	n.a.
	(0.66)	(0.66)	
at wave 5 (n = 473)	0.88	0.88	n.a.
	(0.75)	(0.75)	
Age (centred; in years)			
at wave 1 (n = 556)	0.01	0.01	no miss.
	(0.32)	(0.32)	
at wave 3 (n = 528)	2.00	2.00	0.8%
	(0.31)	(0.31)	
at wave 5 (n = 473)	3.74	3.74	3.8%
	(0.32)	(0.32)	
Education mother			
CASMIN 1a, b, c (low or no education)	11%	11%	8.1%
CASMIN 2a, b (intermediate secondary education with or w/o vocational degree)	41%	41%	
CASMIN 2c_gen, d (university entrance qualification, w/o tertiary degree)	25%	25%	
CASMIN 3a, b (tertiary education)	24%	23%	
Household possessions	0.00	−0.02	12.6%
	(1.00)	(1.02)	
Parental reading (wave 1, 3)	0.00	0.00	5.8%
	(1.00)	(1.01)	
Literacy activities	0.00	0.01	12.6%
	(1.00)	(1.03)	
Reading aloud to child (wave 1, 2, 3)	0.00	−0.01	4.5%
	(1.00)	(1.01)	
Girl = 1, boy = 0	51%	51%	no miss.
Immigration status (at least one grandparent born abroad = 1, else 0)	21%	22%	7.7%
Language at home (non-German = 1, otherwise 0, teacher report)	5%	5%	no miss.
Number of children under age 14 (logarithm)	0.57	0.57	7.7%
	(0.41)	(0.41)	

Notes: All information from wave 1 unless stated otherwise. Abbreviations: SD = standard deviations; n = case numbers; no miss. = no missing values; n.a. = does not apply.

Sources: doi:10.5157/NEPS:SC2:8.0.1; own calculations.

initial skills and progress are independent, the error terms do not correlate. In the case of a positive correlation, achievement inequality increases as predicted, for example by cumulative advantage in strict form. In contrast, a negative correlation indicates that children with initially low skills make more progress (compensatory effects).

To test whether progress differs by family status, we expanded both level-2 equations to include information on the mother's education (EDU). Whereas β_{01} in equation 4 refers to initial differences, β_{11} in equation 5 is a cross-level interaction effect. Inserting equations 4 and 5 into equation 1 leads to equation 6.

$$\text{Level2 (random intercept):}\ \pi_{0i} = \beta_{00} + \beta_{01}EDU_i + \cdots + r_{0i} \tag{4}$$

$$\text{Level2 (random slope):}\ \pi_{1i} = \beta_{10} + \beta_{11}EDU_i + \cdots + r_{1i} \tag{5}$$

$$\begin{aligned} Y_{ti} &= \beta_{00} + \beta_{01}EDU_i + r_{0i} + (\beta_{10} + \beta_{11}EDU_i + r_{1i})Age_{ti} + e_{ti} \\ &= \beta_{00} + \beta_{01}EDU_i + \beta_{10}Age_{ti} + \beta_{11}EDU_iAge_{ti} + e_{ti} + r_{0i} + r_{1i}Age_{ti} \end{aligned} \tag{6}$$

If EDU represents high educational attainment and $\beta_{01} > 0$, $\beta_{10} > 0$, *and* $\beta_{11} > 0$, this indicates an initial advantage as well as an increasing advantage by status over time. Note, however, that it is difficult to directly test the existence of cumulative advantage *in its strict form*. Indications of its existence are present if there is still a positive correlation between both level-2 error terms after controlling for important factors in equation 4 and 5.

Results

Descriptive results

The mean latent value for children's receptive vocabulary skills increases from −0.85 in wave 1 to 0.10 in wave 2 and 0.88 in wave 3. In addition, the standard deviations increase from 0.56 over 0.66 to 0.75 (see Table 1). The same pattern is observable when holding the sample constant by using either imputed values or listwise deletion (see Appendix, Table A1). That is, the increase in variation is not rooted in different sample compositions over time.

The descriptive information on vocabulary for each SES group reveals that children whose mothers have a low or no secondary education lag behind other children at age five and fall even further behind over time (see Table 2). In fact, the average skill level of children with mothers with a low or no secondary education is still somewhat lower at age nine (wave 5, 0.26) than the average skill level of children with highly educated mothers (tertiary education) at age seven (wave 3, 0.34). It is also worth looking at the standard deviations (SD) in Table 2. They increase substantially over time in all subgroups, except for low-SES families. For children whose mothers have a low or no secondary education, the standard deviation increases only slightly (wave 1: 0.69, wave 3: 0.69, wave 5 0.74).

Cumulative advantages in strict form (RQ 1.1) or pure compensatory effects (RQ 2.1)

Cumulative advantages in a strict form (RQ 1.1) and pure compensatory effects (RQ 2.1) are strictly contradictory patterns. To provide evidence of one or the other,

Table 2: Vocabulary skills in the analysis sample by mother's educational attainment: means, standard deviations (SD) in brackets and case numbers (n).

Education	Low or no secondary education (CASMIN 1 a, b, c)		Intermediate secondary education (CASMIN 2 a, b)		Upper secondary education (university entrance qualification) (CASMIN 2 c_gen, d)		Tertiary education (CASMIN 3 a, b)	
Wave	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n
1	−1.20		−0.89		−0.77		−0.70	
	(0.69)	61	(0.53)	226	(0.51)	140	(0.53)	130
3	−0.42		0.05		0.19		0.34	
	(0.69)	59	(0.62)	215	(0.60)	129	(0.60)	125
5	0.26		0.84		1.00		1.11	
	(0.74)	50	(0.73)	192	(0.69)	114	(0.70)	118

Notes: Based on ten fully imputed data sets; SD: standard deviation; n: case numbers.

Sources: doi:10.5157/NEPS:SC2:8.0.1, authors' calculations.

we started with an 'empty' growth curve model (see Table 3, Model 1). Here, only time (age) was included, and a random intercept and a random slope are specified. As expected, the coefficient on age is positive. Each additional year increases the performance level by 0.46. There is not only variation in skills at the age of five years[11] $[0.448 - 1.96 \times 0.02; 0.448 + 1.96 \times 0.02]$ but also in individual-specific progress. The 95% confidence interval of the standard deviation of the error term r_{1i} is $[0.089 - 1.96 \times 0.009; 0.089 + 1.96 \times 0.009]$. Thus, there is considerable variation in progress between children. Variation itself does not say anything about widening or shrinking distribution. The correlation between the error terms at the individual level (r_{0i}, r_{1i}) gives important information on this. This correlation is positive: children with 'initially' above-average (or below-average) performance make more (or less) progress compared to average children over time. This is in line with cumulative advantages in a strict form (RQ 1.1) and clearly contradicts pure compensatory effects (RQ 2.1).

Cumulative advantages by exposure (RQ 1.2)

The first model gives hints of cumulative advantage in strict form. But is this increase only attributable to previous skills, or can this increase in skill variation be partly or completely attributed to family SES? To answer this question, Model 2 includes, in addition to some control variables, the mother's education as the key SES indicator. There is a clear pattern visible at age five: the higher the mother's education, the higher the child's vocabulary skills. The difference between children whose mothers have a low or no education and those with tertiary degrees equals the effect of about one year.[12] With respect to cumulative advantages, the change in gaps is crucial. The gap between children whose mothers fall into the lowest and all other educational groups becomes larger over time. The gaps between the lowest and the two intermediate groups increase by 0.066 and 0.072 per year (p-values at .004 and .000), between the lowest and highest group 0.093 per year (p-value at .000). We calculated the point estimate and 95% confidence interval on the skill level of a boy from a non-immigrant family with German as main language

Table 3: Progress in vocabulary skill from age five to nine: results from growth curve models (Coeff.; (SE), *p*-value)

	Model 1		**Model 2**	
Variables	**Intercept**	**Slope**	**Intercept**	**Slope**
Intercept/slope (age)	−0.842	0.459	−0.979	0.386
	(0.023)	(0.006)	(0.075)	(0.024)
	0.000	0.000	0.000	0.000
Education mother (Ref. CASMIN 1a, b, c = low or no education)				
CASMIN 2a, b (intermediate secondary education with or w/o vocational degree)			0.209	0.066
			(0.073)	(0.023)
			0.004	0.004
CASMIN 2c_gen, d (university entrance qualification w/o tertiary degree)			0.354	0.072
			(0.074)	(0.024)
			0.000	0.003
CASMIN 3a, b (tertiary education)			0.449	0.093
			(0.075)	(0.024)
			0.000	0.000
Girl (ref. boy)			−0.005	0.022
			(0.040)	(0.013)
			0.903	0.084
Immigration status (no)				
At least one grandparent born abroad			−0.152	0.028
			(0.055)	(0.017)
			0.006	0.105
Language at home (teacher report; ref. German)				
Non-German			−0.777	−0.028
			(0.100)	(0.032)
			0.000	0.384
Number of children under age 14 in HH (logarithm)			−0.116	−0.016
			(0.050)	(0.016)
			0.021	0.322
Random effects				
SD error term at level 1	0.305		.306	
(SE)	(0.010)		(.010)	
SD error terms at level 2 (person level)	0.448	0.089	0.362	0.083
(SE)	(0.020)	(0.009)	(0.019)	(0.010)
Correlation of errors at level 2	0.505		0.552	
(SE)	(0.133)		(0.168)	
R^2 at level 2	n.a.	n.a.	34.7%	13.0%
Observations ('person years')	1,557		1,557	
Number of persons	556		556	

Notes: Based on ten fully imputed data sets, Robust standard errors (SE) in parentheses, SD: Standard deviation, HH: household, n.a. = does not apply.

Sources: doi:10.5157/NEPS:SC2:8.0.1; authors' calculations.

in the family and no other child in the household at age nine. In ascending order, the estimates with 95% confidence intervals in brackets are 1.54 [1.36 and 1.73], 2.02 [1.83, 2.20], 2.19 [1.99, 2.38] and finally 2.36 [2.16, 2.56] (not shown in the table). Compared to the estimations at age five (see Model 2, column intercept), the difference is nearly twice as large (2.36–1.54 vs 0.449). This is in line with the perspective of cumulative advantage by exposure.

Note that the differences in growth between the two intermediate groups and the highest group are low. Changing the reference category to mothers with tertiary education reveals *p*-values of .112 and .302 on differences in growth between the three groups (full results on request). Consequently, there is a sharp difference in vocabulary growth between children raised by mothers with a low level of education and all others.

The error terms still correlate highly in Model 2. That is, having an advantageous 'start' leads to more progress, even when controlling for SES and socio-demographic characteristics. This is a hint that cumulative advantage in a strict sense might also be at work.

Compensatory advantage by SES (RQ 2.2)

RQ 2.2 tackles compensatory advantage by SES, which means children form high-SES families are still ahead of other children, but that within this group, they are becoming more homogeneous. To gain insights as to whether the increasing SES gaps already described are accompanied by a process of catching-up of the low-performers in high-SES families, we look at the distribution of the errors, the differences between the observed and predicted vocabulary skills, by mother's education. Figures in Table 4 show some slight increases in the errors across waves in all groups. Thus, there is no indication that children from families with highly educated mothers are becoming more homogeneous in terms of vocabulary skills than children from other families.

Importance of family activities above and beyond SES (RQ 3.1)

In the next step, we addressed RQ 3.1 on the importance of family activities for vocabulary growth above and beyond SES. Table 5 presents different extensions to Model 2 (see Table 3 for Model 2). We include four different indicators of cultural capital and family activities separately, starting with more distal and continuing with more proximate indicators, and present a final model at the end.

The relationship between (a) household possessions and children's vocabulary skills is 0.109 [0.064, 0.154], and the estimated coefficient has a *p*-value of .000 at age five. The coefficient on skill progress is positive as well, 0.013 [−0.002,0.028], with a *p*-value of .086 (see Table 5, Model 3a). In the case of (b) mother's reading time, there is no relationship to the child's skills at age five or to progress in vocabulary (Model 3b), whereas (c) child-related literacy activities in the family correlate positively with vocabulary progress at age 5 (0.048 [0.009, 0.087]). The estimated coefficient has a *p*-value of .016. But these activities are rather unlikely to affect the child's skill progress according to the confidence interval (0.009 [−0.004,0.022], *p*-value = .178) (see Model 3c). One indicator that is often discussed in the literature is reading to the child (for example, Scarborough and Dobrich, 1994; Mol et al, 2008;

Table 4: Standard deviations (SD) of residuals (observed vs fitted values) based on Model 2, Table 3, and case numbers (n).

Education	Low or no secondary education (CASMIN 1 a, b, c)		Intermediate secondary education (CASMIN 2 a, b)		Upper secondary education (university entrance qualification) (CASMIN 2 c_gen, d)		Tertiary education (CASMIN 3 a, b)	
Wave	SD	n	SD	n	SD	n	SD	n
1	0.55	61	0.45	226	0.44	140	0.49	130
3	0.60	59	0.58	215	0.53	129	0.58	125
5	0.65	50	0.70	192	0.62	114	0.67	118

Notes: Based on ten fully imputed data sets; SD: standard deviation; n: case numbers.

Sources: doi:10.5157/NEPS:SC2:8.0.1, authors' calculations.

Sénéchal et al, 2008; Klein and Kogan, 2013). Therefore, we use the information collected in waves 1, 2 and 3 (Model 3d). We find positive relationships to the 'initial' skill level and to progress in vocabulary, with estimated 0.061 [0.021,0.101], *p*-value of .003, and 0.016 [0.003, 0.029] (*p*-value of .017).

In Models 3a to 3d, household possessions and reading to child seem to have the most influence. Compared to the mother's time devoted to reading, household possessions appear to be fairly stable characteristics. Mothers' time reading might be more volatile depending on current family and work circumstances. With regard to activities with the child, the indicator most closely related to written language shows somewhat stronger correlations with language development than the composite score. This is not astonishing, as activities such as looking at picture books together might have little impact on the child's language skills if the parent's own language skills are relatively limited. Written language offers a wider range of vocabulary and more complex sentences than everyday language, independent of the parent's skills.

In the final model, we included household possessions, which indicate cultural capital in its objective form, and reading to the child (see Model 3e). With respect to the initial skill level, both estimated coefficients indicate that more household possessions and reading more often to the child go hand in hand with higher vocabulary skills at age five. Concerning skill progress, both coefficients are positive, but looking at both confidence intervals [−0.006,0.025] and [0.000,0.027] reading to the child is slightly more associated with improvements in children's vocabulary skills. With respect to RQ 3.1, we conclude that although three out of four indicators positively correlate with the skill level at age five, when it comes to predicting changes over time, parental reading seems to be the strongest indicator in this age range of the indicators considered here.

Importance of family activities in mediating SES effects on vocabulary growth (RQ 3.2)

RQ 3.2 addresses the potential mechanisms behind changing SES gaps in vocabulary skills and goes deeper into testing cumulative advantages by exposure. We explore this question by looking at changes in the estimated coefficient on the mother's education in Model 3e.

At age five, the estimated coefficients on mother's education decrease by around 30% to 37%. With respect to SES differences in progress, the effects of the mother's

Table 5: Progress in vocabulary skills from age five to nine. Table 3, Model 2, extended by covariates on cultural capital and activities (Coeff.; (SE), *p*-value)

	Model 3a		**Model 3b**		**Model 3c**		**Model 3d**		**Model 3e**	
Variables	Int.	Slope	Int.	Slope	Int.	Slope	Int.	Slope	Int.	Slope
Intercept/slope (age)	−0.870	0.399	−0.978	0.386	−0.966	0.388	−0.960	0.390	−0.869	0.399
	(0.075)	(0.025)	(0.075)	(0.024)	(0.074)	(0.024)	(0.074)	(0.024)	(0.075)	(0.024)
	0.000	0.000	0.000	0.000	0.000	0.000	0.00	0.000	0.000	0.000
Education mother (ref. low)										
CASMIN 2a, b	0.153	0.059	0.210	0.066	0.199	0.064	0.190	0.061	0.147	0.057
	(0.071)	(0.023)	(0.073)	(0.023)	(0.073)	(0.023)	(0.072)	(0.023)	(0.071)	(0.023)
	0.032	0.010	0.004	0.004	0.006	0.005	0.009	0.007	0.040	0.013
CASMIN 2c_gen, d	0.238	0.058	0.354	0.072	0.351	0.072	0.334	0.068	0.237	0.059
	(0.076)	(0.026)	(0.074)	(0.024)	(0.074)	(0.024)	(0.074)	(0.024)	(0.075)	(0.026)
	0.002	0.024	0.000	0.003	0.000	0.003	0.000	0.005	0.002	0.022
CASMIN 3a, b	0.293	0.074	0.449	0.093	0.440	0.091	0.409	0.082	0.284	0.070
	(0.079)	(0.027)	(0.075)	(0.024)	(0.074)	(0.024)	(0.075)	(0.025)	(0.079)	(0.027)
	0.000	0.006	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.008
Household possessions	0.109	0.013							0.098	0.009
	(0.023)	(0.008)							(0.024)	(0.008)
	0.000	0.086							0.000	0.251
Parental reading			−0.007	−0.000						
			(0.020)	(0.007)						
			0.718	0.987						
Literacy activities					0.048	0.009				
					(0.020)	(0.007)				
					0.016	0.178				

(*Continued*)

Table 5: (*Continued*)

	Model 3a		**Model 3b**		**Model 3c**		**Model 3d**		**Model 3e**	
Reading to child							0.061	0.016	0.039	0.014
							(0.021)	(0.007)	(0.021)	(0.007)
							0.003	0.017	0.064	0.047
Random effects										
SD random error level 1	0.306		0.306		0.306		0.306		0.306	
(SE)	(0.010)		(0.010)		(0.010)		(0.010)		(0.010)	
SD error level 2	0.350	0.083	0. 362	0.083	0. 359	0.083	0. 357	0.082	0. 348	0.082
(SE)	(0.019)	(0.010)	(0.019)	(0.010)	(0.019)	(0.010)	(0.019)	(0.010)	(0.019)	(0.010)
Corr. level 2	0.534		0.552		0.548		0.535		0.524	
(SE)	(0.170)		(0.168)		(0.169)		(0.170)		(0.171)	
R^2 at level 2	39.0%	13.0%	34.7%	13.0%	35.8%	13.0%	36.5%	15.1%	39.7%	15.1%
Observations ('person years')	1,557		1,557		1,557		1,557		1,557	
Number of persons	556		556		556		556		556	

Notes: Based on ten fully imputed data sets, Robust standard errors (SE) in parentheses, SD: Standard deviation. Controlling for gender, immigration status, language at home and number of children in household under age 14.

Sources: doi:10.5157/NEPS:SC2:8.0.1; authors' calculations.

education are reduced by 14% to 25% compared to Table 3, Model 2. Overall, household possessions and reading to children 'explain' some variation in children's initial skill differences and at least partly mediate the relationship between SES and initial skills. In comparison, their influence on skill progress is weaker.

We would like to stress that the correlation of the error terms at level 2 of Model 3e in Table 4 is positive, large and comparable in size to all previous models. Even if a family's cultural capital and activities are taken into account, there are hints of cumulative advantage in its strict form. As noted at the end of the Methods section, it is difficult to directly test the existence of cumulative advantage in its strict form, and the reported correlation is only an indication. There might be other factors, such as general cognitive skills of the child or unobserved family or environmental characteristics that are responsible for better test results at age five and more progress in the following years (for a critique of past research that has stressed the importance of early skills for later skill development, see Bailey et al, 2018).[13]

Sensitivity checks

We ran several sensitivity checks. First, we expanded Model 2, Table 3, as well as Table 5, Model 3e, to include the father's education to get a broader picture of family SES. If a father has a higher level of education than the mother, children show somewhat higher vocabulary skills at age five compared to parents who have similar educational levels. This gap does not increase over time. All other constellations, 'father has lower education than mother' and 'mother is a single parent', do not show any further difference compared to parents with educational homogamy (see Appendix Table A2, Model A1, and Table A3, Model A4). The main results presented in Tables 3 and 5 still hold if information on the father's education is included. The higher mother's education, the higher the child's test performance at age five, and these differences increase over time. In Table A3, Model A4, the estimated coefficients on household possessions and reading aloud to the child are comparable to those reported in Table 4, Model 3e. We also estimated a model with a squared term for age, as the correlations between age and test results become weaker over the measurement points (see Section Analytical procedure: growth curve model). Results indicate that the increase in vocabulary skills decelerates with age. Nevertheless, the main findings remain stable (see Appendix Table A2, Model A2; Table A3, Model A5). Finally, as children are clustered within preschools in this study, we also introduced a third level for preschools to the growth curve model. The results did not change (compare Table 3, Model 2, with Appendix Table A2, Model A3; and Table 4, Model 3e, with Appendix Table A3, Model A6).

Conclusions

Early language skills are considered a key factor for educational success, as there is abundant evidence on close links to literacy and reading but also to measures of more domain-general cognitive skills. Looking at research on children's vocabulary skills, both individual differences and early SES gaps have been reported. Family resources and linguistically stimulating activities have been shown to partially explain the differences. Highlighting the importance of academic skills at school entry for later educational outcomes, the openness of vocabulary to environmental influence and

the existence of striking socio-economic differences, Shonkoff and Phillips (2000: 125) also underlined that there is a substantial amount of variance unexplained which leaves room to defy the odds. We looked at the variance in vocabulary development from age five to nine and focused on how pathways depend on socio-economic background and more immediate environmental influences during this period.

Sociology offers different and partly conflicting theories for the explanation of socially unequal developmental trajectories. We highlighted cumulative advantages (so-called Matthew effects) and compensatory advantages. We differentiated 'strict' and 'pure' types, where the amount of initial skills is the main driver for subsequent increases or decreases in disparities, from those types, where rather exogenous factors such as SES indicators are expected to drive the development of initial inequalities.

Differentiating these mechanisms and also considering explanatory factors introduced by Bourdieu (1984), Lareau (2011), and research on language development, we asked: (1.1) Are there cumulative advantages in their strict form? (1.2) Do we see increasing SES gaps in vocabulary skills even if we control for potential path dependency? (2.1) Do we find pure compensatory effects or (2.2) advantages by parents' SES? And (3.1) how important are parents' cultural capital and activities as well as parent–child activities for children's progress in vocabulary above and beyond SES and (3.2) for 'explaining' the relationship between SES and vocabulary?

As longitudinal research on the development of early inequalities in vocabulary from kindergarten into the first years of school is rather sparse, especially when evaluating competing mechanisms, we used recent nationwide data from Germany to analyse the development of social inequalities from age five to nine and used growth curve models for our analyses, as they allow differentiation between effects that root back to initial skills and those that determine progress in skills.

As expected, vocabulary skills increase from age five to nine in Germany. There are clear hints that the distribution of vocabulary skills increases over time and that those with above-average skills at age five will have even higher skills at age nine. As this pattern is consistent across all models, which include several structural indicators and some family activities, this supports the assumption that cumulative advantage in the strict form is at work in this age range (RQ 1.1). As stated at the end of the sections 'Analytical Procedure' and 'Importance of Family Activities in Mediating SES Effects on Vocabulary Growth', we cannot fully rule out the possibility that other unobserved factors are responsible for this correlation, and therefore we have to be cautious.

Additionally, SES gaps in these skills become more pronounced over time. This is partly due to the fact that children whose mothers have a low or no education show far below-average increases in their language skills. At the age of five, these children are, on average, one year behind children from families in which mothers have a tertiary degree and this gap doubles in the four years that follow. These findings are clearly in line with cumulative advantage due to 'exposure' (RQ 1.2). There are no signs of general compensatory effects (RQ 2.1) as the gaps in skills do not weaken as children grow older. In contrast: the skill distribution increases over time. Further on, there are no hints on compensatory advantages by status (RQ 2.2), which predicts that children from upper-SES families who initially perform worse would make comparatively greater progress.

The use of different indicators of cultural capital and activities reveals that vocabulary skills at age five can be at least partly attributed to household possessions as well as

activities at home, above all reading to the child. With respect to progress, among the indicators used in this study, reading to the child, especially, seems to be important. Parents' own reading time did not turn out to be relevant for vocabulary development during this age span above and beyond SES (RQ 3.1). The last research question (3.2) asked whether these household characteristics and family activities mediate SES gaps in skills. Even when controlling for several indicators, especially reading to the child, there are still substantial correlations between SES and the child's 'initial' skills at age five, as well as to the child's progress. At the same time, partial mediation was observed, in which about 30% of the direct relationship between SES and vocabulary skills could be explained by the covariates.

The finding that SES gaps in vocabulary skills increase might be surprising, as all children attend preschool and later on elementary school, where they share the same curriculum. However, this result is in line with previous studies on vocabulary development in Germany that focused on an even earlier age range in early childcare (Schmerse et al, 2018) as well as studies carried out in the German capital on a later age range in elementary school (Kigel et al, 2015). Furthermore, most studies that estimate the influence of school on SES gaps by comparing children's development when school is in session versus out of session show that school plays only a modest role in explaining achievement gaps (see, for example, Downey et al, 2017).

The findings on increasing gaps overall and by status seem to be in line with Bourdieu's cultural capital approach and Lareau's concept of concerted cultivation. However, the lowest-SES group clearly differs most from the other SES groups. This contrasts with the idea that especially a highly privileged 'dominant class' (Bourdieu, 1984; 1986) or (upper) middle class (Lareau, 2011) differs and sets itself apart from the others.

In addition, the rather weak relationship between family activities and children's language skills might be surprising. However, in the cross-sectional part of our model focusing on skills at age five, the correlation is (in most cases) stronger than in the part focusing on progress over time. Finally, even some studies based on large samples show a relatively modest or no impact of such indicators when shifting from a cross-sectional to a longitudinal perspective (Taylor et al, 2013; McMullin et al, 2020: 618).

A clear limitation is that our analysis starts at age five, where we already see stark differences in language skills by SES, which might be the result of family activities in general and SES-specific differences in these activities. Information that is currently lacking in social stratification research includes the language skills of parents and other household members as well as the specific financial and time investment of parents in the first years of children's lives and from age five on. A further limitation is the low sample size, which causes rather large sampling errors. Therefore, we might have been to quick ruling out some family activities as important and the results are rather conservative.

Notes

[1] This effect takes its name from the book of Matthew in the Bible: 'To all those who have, more will be given, and they will have an abundance; but from those who have nothing, even what they have will be taken away' (Matthew 13:12 [see also 25:29], *The New Oxford Annotated Bible*, New Revised Standard Version).

[2] For an application of DiPrete and Eirich's classification in educational research, see Baumert et al (2012).

[3] With regard to students' activities, the causal relationship has to be addressed, as children with higher language proficiencies might self-sort into reading: 'Children who become better readers have selected […], shaped […], and evoked […] an environment that will be conductive to further growth in reading' (Stanovich, 1986: 382).

[4] The factor comprised parents' school involvement, children's participation in organised (learning-conducive) leisure activities, and the number of books at home.

[5] This paper uses data from the second starting cohort of the National Educational Panel Study (NEPS), doi:10.5157/NEPS:SC2:8.0.1. From 2008 to 2013, NEPS data were collected as part of the Framework Programme for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). Since 2014, NEPS has been carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network.

[6] The share of parents with an immigration status who provided at least one interview in wave 3 or 5 in addition to their interview in wave 1 is 32%. The low share of children from immigrant families in the longitudinal subsample is only partly due to panel drop-out. One reason for the loss of children from immigrant families in the longitudinal subsample might be that they more often live in cities where there is a wider range of preschools and elementary schools available and therefore a looser link between preschools and elementary schools.

[7] Twenty-three items were administered in all waves; a further 15 identical items were administered in waves 1 and 3; a further 4 items were identical in waves 1 and 5; and a further 25 items were identical in waves 3 and 5.

[8] The first item in wave 1 had to be discarded as all children in our sample solved it. One item from wave 5 is currently not in the scientific use file for unknown reasons.

[9] About 93% of the interviewees are mothers, the rest fathers.

[10] Auxiliary variables used for imputations are: parents' participation in highbrow and pop culture, father's education, welfare receipt, satisfaction with financial situation, single parenthood (parent and preschool report), child's age at enrolment in early childcare, hours per week in preschool, child's test results in science and general cognitive functioning (speed and reasoning, assessed in wave 2), a scale on child's participation in sports, music, language or other course in leisure time (asked in wave 2), and the following information provided by preschool staff: rating of child's skills in understanding German compared to children of the same age, child's need for language remediation, a scale on children's activities in preschool, subscale of the Strengths and Difficulties Questionnaire (SDQ) on prosocial behaviour and problem behaviour as well as disruptive behaviour (TASB).

[11] Remember that the age of five years was set to 0.

[12] The coefficient for CASMIN 3a, 3b is 0.449 in Model 2, the coefficient on age is 0.459 in Model 1, in Model 2 the effect on progress by age is 0.386 in the case of mothers with low education and 0.386 + 0.093 = 0.479 in the case of mothers with tertiary education.

[13] As a robustness check, we estimated models with child's result in a speed and a reasoning test conducted in wave 2. Our main results do not change (results on request).

Acknowledgements

We thank Jane Waldfogel for her very helpful comments, which contributed to a substantial improvement of the manuscript. All errors are our own.

Data availability statement

We are responsible for data preparation and model specification. All syntax files are available on request. We use data from the German National Educational Panel Study (NEPS). The scientific use files of NEPS are easy to access by the scientific community. Experimentation on humans and animals statement: does not apply.

Conflict of interest

The authors declare that there is no conflict of interest.

References

Alexander, K.L., Entwisle, D.R. and Olson, L.S. (2007) Lasting consequences of the summer learning gap, *American Sociological Review*, 72(2): 167–80. doi: 10.1177/000312240707200202

Bailey, D.H., Duncan, G.J., Watts, T., Clements, D.H. and Sarama, J. (2018) Risky business: correlation and causation in longitudinal studies of skill development, *American Psychologist*, 73(1): 81–94. doi: 10.1037/amp0000146

Baumert, J., Nagy, G. and Lehmann, R. (2012) Cumulative advantages and the emergence of social and ethnic inequality: matthew effects in reading and mathematics development within elementary schools?, *Child Development*, 83(4): 1347–67. doi: 10.1111/j.1467-8624.2012.01779.x

Berendes, K., Weinert, S., Zimmermann, S. and Artelt, C. (2013) Assessing language indicators across the lifespan within the German National Educational Panel Study (NEPS), *Journal for Educational Research Online / Journal für Bildungsforschung Online*, 5(2): 15–49.

Bernardi, F. (2014) Compensatory advantage as a mechanism of educational inequality: a regression discontinuity based on month of birth, *Sociology of Education*, 87(2): 74–88. doi: 10.1177/0038040714524258

Bernardi, F. and Grätz, M. (2015) Making up for an unlucky month of birth in school: causal evidence on the compensatory advantage of family background in England, *Sociological Science*, 2(12): 235–51. doi: 10.15195/v2.a12

Blossfeld, H.P. and Roßbach, H.G. (eds) (2019) *Education as a Lifelong Process: The German National Educational Panel Study (NEPS)*, 2nd edn, Wiesbaden: VS Verlag für Sozialwissenschaften (Zeitschrift für Erziehungswissenschaft. Sonderheft Bd. 14).

Bourdieu, P. (1984) *Distinction: A Social Critique of the Judgement of Taste*, Cambridge, MA: Harvard University Press.

Bourdieu, P. (1986) The forms of capital, in J. Richardson (ed) *Handbook of Theory and Research for the Sociology of Education*, Westport, CT: Greenwood, pp 241–58.

Bradbury, B., Corak, M., Waldfogel, J. and Washbrook, E. (2015) *Too Many Children Left Behind: The US Achievement Gap in Comparative Perspective*, New York: Russell Sage Foundation.

Braun, M. and Müller, W. (1997) Measurement of education in comparative research, *Comparative Social Research*, 16: 163–201.

Bronfenbrenner, U. and Morris, P.A. (2006) The bioecological model of human development, in R.M. Lerner (ed) *Handbook of Child Psychology. Volume One: Theoretical Models of Human Development*, 6th edn, Hoboken, NJ: John Wiley, pp 793–828.

Bukodi, E. and Goldthorpe, J.H. (2013) Decomposing 'social origins': the effects of parents' class, status, and education on the educational attainment of their children, *European Sociological Review*, 29(5): 1024–39. doi: 10.1093/esr/jcs079

Bynner, J. (2004) *Participation and Progression: Use of Birth Cohort Study Data in Illuminating the Role of Basic Skills and Other Factors*, Nuffield Review of 14–19 Education and Training Working Paper 9, Oxford: Nuffield.

Cheadle, J.E. (2008) Educational investment, family context, and children's math and reading growth from kindergarten through the third grade, *Sociology of Education*, 81(1): 1–31. doi: 10.1177/003804070808100101

Cunha, F., Heckman, J.J., Lochner, L. and Masterov, D.V. (2006) Interpreting the evidence on life cycle skill formation, in E. Hanushek and F. Welch (eds) *Handbook of the Economics of Education*, Vol. 1, 2nd edn, Amsterdam: Elsevier, pp 698–812.

Curran, P.J., Obeidat, K. and Losardo, D. (2010) Twelve frequently asked questions about growth curve modeling, *Journal of Cognition and Development*, 11(2): 121–36. doi: 10.1080/15248371003699969

Davis-Kean, P.E. (2005) The influence of parent education and family income on child achievement: the indirect role of parental expectations and the home environment, *Journal of Family Psychology*, 19(2): 294-304.

DiPrete, T.A. and Eirich, G.M. (2006) Cumulative advantage as a mechanism for inequality: a review of theoretical and empirical developments, *Annual Review of Sociology*, 32: 271–97. doi: 10.1146/annurev.soc.32.061604.123127

Downey, D.B. (2001) Number of siblings and intellectual development: the resource dilution explanation, *American Psychologist*, 56(6/7): 497–504. doi: 10.1037/0003-066X.56.6-7.497

Downey, D.B., Workman, J. and von Hippel, P. (2017) Socioeconomic, racial, and gender gaps in children's social/behavioral skills: do they grow faster in school or out?, *Sociological Science*, 6(17): 446–66. doi: 10.15195/v6.a17

Duncan, G.J. and Magnuson, K.A. (2003) Off with Hollingshead: socioeconomic resources, parenting, and child development, in M.H. Bornstein and R.H. Bradley (eds) *Socioeconomic Status, Parenting, and Child Development*, Mahwah, NJ: Lawrence Erlbaum, pp 83–106.

Dunn, L.M. and Dunn, D.M. (2004) *Peabody Picture Vocabulary Test (PPVT) (German version)*, Göttingen: Hogrefe.

Duzy, D., Souvignier, E., Ehm, J.H. and Gold, A. (2014) Early decoding speed and later reading competencies in children with German as a second language, *Child Indicators Research*, 7(4): 787–804. doi: 10.1007/s12187-014-9242-x

Ebert, S. (2011) *Was Kinder über die Mentale Welt Wissen-die Entwicklung von Deklarativem Metagedächtnis aus der Sicht der 'Theory of Mind' [What Children Know About the Mental World: The Development of Declarative Metamemory from the Perspective of 'Theory of Mind']*, Hamburg: Dr. Kovač.

Feinstein, L. (2003) Inequality in the early cognitive development of British children in the 1970 cohort, *Economica*, 70(277): 73–97. doi: 10.1111/1468-0335.t01-1-00272

Feinstein, L., Jerrim, J., Vignoles, A., Goldstein, H., French, R., Washbrook, E., Lee, R. and Lupton, R. (2015) Comment and debate: social class differences in early cognitive development, *Longitudinal and Life Course Studies*, 6(3): 331–76.

Fernald, A., Marchman, V.A. and Weisleder, A. (2013) SES differences in language processing skill and vocabulary are evident at 18 months, *Developmental Science*, 16(2): 234–48. doi: 10.1111/desc.12019

Garcia, E. (2015) *Inequalities at the Starting Gate: Cognitive and Noncognitive Skills Gaps Between 2010–2011 Kindergarten Classmates*, Washington, DC: Economic Policy Institute.

Hart, B. and Risley, T.R. (1995) *Meaningful Differences in the Everyday Experience of Young American Children*, Baltimore, MD: Paul H. Brookes Publishing.

Heckman, J.J. and Masterov, D.V. (2007) The productivity argument for investing in young children, *Review of Agricultural Economics*, 29(3): 446–93. doi: 10.1111/j.1467-9353.2007.00359.x

Hoff, E. (2003) The specificity of environmental influence: socioeconomic status affects early vocabulary development via maternal speech, *Child Development*, 74(5): 1368–78. doi: 10.1111/1467-8624.00612

Hoff, E. (2006a) Environmental supports for language acquisition, in D.K. Dickinson and S.B. Neuman (eds) *Handbook of Early Literacy Research*, Vol. 2, New York: Guilford Press, pp 163–72.

Hoff, E. (2006b) How social contexts support and shape language development, *Developmental Review*, 26(1): 55–88. doi: 10.1016/j.dr.2005.11.002

Jerrim, J. and Vignoles, A. (2013) Social mobility, regression to the mean and the cognitive development of high ability children from disadvantaged homes, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(4): 887–906. doi: 10.1111/j.1467-985X.2012.01072.x

Kerckhoff, A.C., Ezell, E.D. and Brown, J.S. (2002) Toward an improved measure of educational attainment in social stratification research, *Social Science Research*, 31(1): 99–123. doi: 10.1006/ssre.2001.0720

Kigel, R.M., McElvany, N. and Becker, M. (2015) Effects of immigrant background on text comprehension, vocabulary, and reading motivation: a longitudinal study, *Learning and Instruction*, 35: 73–84. doi: 10.1016/j.learninstruc.2014.10.001

Klein, O. and Kogan, I. (2013) Does reading to children enhance their educational success?, *Child Indicators Research*, 6(2): 321–44. doi: 10.1007/s12187-012-9174-2

Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M., Schneider, W. and Stanat, P. (eds) (2010) *PISA 2009: Bilanz Nach Einem Jahrzehnt*, Münster: Waxmann.

Kristen, C., Olczyk, M. and Will, G. (2016) Identifying immigrants and their descendants in the National Educational Panel Study, in H.P. Blossfeld, J. von Maurice, M. Bayer and J. Skopek (eds) *Methodological Issues of Longitudinal Surveys: The Example of the National Educational Panel Study*, Wiesbaden: Springer, pp 195–211.

Lareau, A. (2011) *Unequal Childhoods. Class, Race, and Family Life*, 2nd edn, Berkeley, CA: University of California Press.

Lee, V.E. and Burkam, D.T. (2002) *Inequality at the Starting Gate: Social Background Differences in Achievement as Children Begin School*, Washington, DC: Economic Policy Institute.

Linberg, T. and Wenz, S.E. (2017) Extent and distribution of socio-economic and Immigration-related inequalities in the language abilities of 5-year-old preschool children, *Journal for Educational Research Online / Journal für Bildungsforschung Online*, 9(1): 77–98.

Linberg, T., Schneider, T., Waldfogel, J. and Wang, Y. (2019) Socioeconomic status gaps in child cognitive development in Germany and the United States, *Social Science Research*, 79: 1–31. doi: 10.1016/j.ssresearch.2018.11.002

Magnuson, K.A. and Duncan, G.J. (2006) The role of family socioeconomic resources in the Black–white test score gap among young children, *Developmental Review*, 26(4): 365–99. doi: 10.1016/j.dr.2006.06.004

McMullin, P., McGinnity, F., Murray, A. and Russell, H. (2020) What you do versus who you are: home-learning activities, social origin and cognitive skills among young children in Ireland, *European Sociological Review*, 36(4): 610–25. doi: 10.1093/esr/jcaa012

Merton, R.K. (1968) The Matthew effect in science, *Science*, 159(3810): 56–63. doi: 10.1126/science.159.3810.56

Merton, R.K. (1988) The Matthew effect in science, II: cumulative advantage and the symbolism of intellectual property, *Isis*, 79(4): 606–23. doi: 10.1086/354848

Mol, S.E., Bus, A.G., de Jong, M.T. and Smeets, D.J.H. (2008) Added value of dialogic Parent–child book readings: a meta-analysis, *Early Education and Development*, 19(1): 7–26. doi: 10.1080/10409280701838603

OECD (2010) *PISA 2009 Results: Overcoming Social Background – Equity in Learning Opportunities and Outcomes (Volume II)*, Paris: OECD.

Pellegrini, A.D. (2002) Some theoretical and methodological considerations in studying literacy in social context, in S. Neuman and D. Dickinson (eds) *Handbook of Early Literacy Research*, Vol. 1, New York: Guilford Press, pp 54–65.

Pensiero, N. (2011) Parent–child cultivation and children's cognitive and attitudinal outcomes from a longitudinal perspective, *Child Indicators Research*, 4(3): 413–37. doi: 10.1007/s12187-011-9106-6

Pfost, M., Hattie, J., Dörfler, T. and Artelt, C. (2014) Individual differences in reading development: a review of 25 years of empirical research on Matthew effects in reading, *Review of Educational Research*, 84(2): 203–44. doi: 10.3102/0034654313509492

Quinn, D.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. doi: 10.1177/0038040715573027

Raudenbush, S.W. and Bryk, A. (2002) *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd edn, Thousand Oaks, CA: Sage.

Raudenbush, S.W. and Eschmann, R.D. (2015) Does schooling increase or reduce social inequality?, *Annual Review of Sociology*, 41: 443–70. doi: 10.1146/annurev-soc-071913-043406

Rice, M.L. and Hoffman, L. (2015) Predicting vocabulary growth in children with and without specific language impairment: a longitudinal study from 2;6 to 21 years of age, *Journal of Speech, Language, and Hearing Research*, 58(2): 345–59. doi: 10.1044/2015_JSLHR-L-14-0150

Rowe, M.L., Raudenbush, S.W and Goldin-Meadow, S. (2012) The pace of vocabulary growth helps predict later vocabulary skill, *Child Development*, 83(2): 508–25. doi: 10.1111/j.1467-8624.2011.01710.x

Rubin, D.B. (2004) *Multiple Imputation for Nonresponse in Surveys*, Hoboken, NJ: John Wiley.

Sameroff, A.J. (2010) A unified theory of development: a dialectic integration of nature and nurture, *Child Development*, 81(1): 6–22. doi: 10.1111/j.1467-8624.2009.01378.x

Sammons, P., Sylva, K., Melhuish, E., Siraj-Blatchford, I. and Taggart, B. (2007) *Effective Pre-school and Primary Education 3–11 Project (EPPE 3–11): Summary Report, Influences on Children's Attainment and Progress in Key Stage 2: Cognitive Outcomes in Year 5*, London: Institute of Education, University of London/Department for Education and Skills.

Scarborough, H.S. and Dobrich, W. (1994) On the efficacy of reading to preschoolers, *Developmental Review*, 14(3): 245–302, doi: doi: 10.1006/drev.1994.1010.

Schmerse, D. Anders, Y., Flöter, M., Wieduwilt, N., Roßbach, H.G. and Tietze, W. (2018) Differential effects of home and preschool learning environments on early language development, *British Educational Research Journal*, 44(2): 338–57. doi: 10.1002/berj.3332

Sénéchal, M., Pagan, S., Lever, R. and Ouellette, G.P. (2008) Relations among the frequency of shared reading and 4-year-old children's vocabulary, morphological and syntax comprehension, and narrative skills, *Early Education and Development*, 19(1): 27–44.

Shaywitz, B.A., Holford, T.R., Holahan, J.M., Fletcher, J.M., Stuebing, K.K., Francis, D.J. and Shaywitz, S.E. (1995) A Matthew effect for IQ but not for reading: results from a longitudinal study, *Reading Research Quarterly*, 30(4): 894–906. doi: 10.2307/748203

Shonkoff, J.P. and Phillips, D.A. (eds) (2000) *From Neurons to Neighborhoods: The Science of Early Childhood Development*, Washington, DC: National Academies Press.

Stanovich, K.E. (1986) Matthew effects in reading: some consequences of individual differences in the acquisition of literacy, *Reading Research Quarterly*, 21(4): 360–407. doi: 10.1598/RRQ.21.4.1

Steinhauer, H.W., Zinn, S., Gaasch, C. and Goßmann, S. (2016) *NEPS Technical Report for Weighting: Weighting the Sample of Kindergarten Children and Grade 1 Students of the National Educational Panel Study (Wave 1 to 3)*, NEPS Working Paper 66, Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study.

Taylor, C.L. Christensen, D., Lawrence, D., Mitrou, F. and Zubrick, S.R. (2013) Risk factors for children's receptive vocabulary development from four to eight years in the Longitudinal Study of Australian Children, *PLOS ONE*, 8(9): 1–20, https://doi.org/10.1371/journal.pone.0073046.

Vasilyeva, M. and Waterfall, H. (2011) Variability in language development: relation to socioeconomic status and environmental input, in S. Neuman (ed) *Handbook of Early Literacy Research*, Vol. 3, New York: Guilford Press, pp 36–48.

Verhoeven, L., van Leeuwe, J. and Vermeer, A. (2011) Vocabulary growth and reading development across the elementary school years, *Scientific Studies of Reading*, 15(1): 8–25. doi: 10.1080/10888438.2011.536125

Walberg, H.J. and Tsai, S.L. (1983) Matthew effects in education, *American Educational Research Journal*, 20(3): 359–73.

Weinert, S. (2010) Beziehungen zwischen Sprachentwicklung und Gedächtnisentwicklung, in H.P. Trolldenier, W. Lenhard and P. Marx (eds) *Brennpunkte der Gedächtnisforschung: Entwicklungs- und pädagogisch-psychologische Perspektiven*, Göttingen: Hogrefe, pp 147–70.

Weinert, S. and Ebert, S. (2013) Spracherwerb im Vorschulalter: soziale Disparitäten und Einflussvariablen auf den Grammatikerwerb, *Zeitschrift für Erziehungswissenschaft*, 16(2): 303–32.

Appendix

Table A1: Vocabulary skills in the analysis sample, in the case of multiple imputed skills and in a complete case sample – mean, standard deviation (in brackets) and case numbers (n)

	Without imputation		With imputation (m = 10)		Complete case	
Wave		**n**		**n**		**n**
1	−0.85	566	−0.85	566	−0.82	445
	(0.56)		(0.56)		(0.56)	
3	0.10	528	0.10	566	0.13	445
	(0.66)		(0.66)		(0.66)	
5	0.88	473	0.86	566	0.88	445
	(0.75)		(0.76)		(0.76)	

Sources: doi:10.5157/NEPS:SC2:8.0.1; authors' calculations.

Table A2: Additional growth curve models on vocabulary skills – extensions of Table 3, Model 2

	Model A1		**Model A2**			**Model A3**		
Variables	**Intercept**	**Slope**	**Intercept**	**Slope**	**Slope^2**	**Intercept**	**Slope**	**Slope^2**
Intercept/slope (age)	−1.045	0.388	−0.997	0.440	−0.015	−0.995	0.440	−0.014
	(0.079)	(0.025)	(0.075)	(0.028)	(0.004)	(0.074)	(0.028)	(0.004)
	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.001
Education mother (ref. CASMIN 1 a, b, c)								
CASMIN 2a, b	0.234	0.068	0.212	0.065		0.206	0.065	
	(0.074)	(0.023)	(0.073)	(0.022)		(0.072)	(0.022)	
	0.002	0.003	0.004	0.004		0.004	0.004	
CASMIN 2c_gen, d	0.366	0.076	0.356	0.071		0.351	0.071	
	(0.076)	(0.025)	(0.074)	(0.024)		(0.074)	(0.024)	
	0.000	0.003	0.000	0.004		0.000	0.004	
CASMIN 3a, b	0.513	0.095	0.452	0.092		0.440	0.092	
	(0.078)	(0.026)	(0.075)	(0.024)		(0.075)	(0.024)	
	0.000	0.003	0.000	0.000		0.000	0.000	
Education father (ref. same level as mother)								
higher than mother	0.171	0.002						
	(0.055)	(0.018)						
	0.002	0.925						
lower than mother	0.021	−0.009						
	(0.052)	(0.017)						
	0.689	0.597						
no father (= mother lone parent)	0.061	−0.010						
	(0.078)	(0.027)						
	0.433	0.713						
Girl (ref. boy)	−0.001	0.022	−0.000	0.019		−0.001	0.019	
	(0.039)	(0.013)	(0.039)	(0.013)		(0.039)	(0.013)	
	0.970	0.085	0.992	0.126		0.972	0.127	
Immigration status (no)								
At least one grandparent born abroad	−0.154	0.027	−0.153	0.029		−0.149	0.029	
	(0.055)	(0.017)	(0.055)	(0.017)		(0.055)	(0.017)	
	0.005	0.115	0.005	0.093		0.007	0.092	
Language at home (teacher report; ref. German)								
Non-German	−0.786	−0.027	−0.775	−0.030		−0.776	−0.030	
	(0.100)	(0.032)	(0.100)	(0.032)		(0.100)	(0.032)	
	0.000	0.407	0.000	0.353		0.000	0.351	
Number of children under age 14 in HH (logarithm)	−0.128	−0.018	−0.115	−0.016		−0.112	−0.016	
	(0.051)	(0.017)	(0.050)	(0.016)		(0.051)	(0.016)	
	0.012	0.277	0.023	0.314		0.027	0.313	
Random effects								
SD random error term at level 1	0.306		0.305			0.305		

(*Continued*)

Table A2: (*Continued*)

	Model A1		Model A2			Model A3		
Variables	**Intercept**	**Slope**	**Intercept**	**Slope**	**Slope^2**	**Intercept**	**Slope**	**Slope^2**
(SE)	(0.010)		(0.010)			(.010)		
SD error terms at level 2 (person level)	0.357	0.083	0.362	0.083	fix	0.352	0.083	fix
(SE)	(0.019)	(0.010)	(0.019)	(0.010)	fix	(0.021)	(0.010)	fix
Correlation of errors at level 2	0.555		0.565			0.573		
(SE)	(0.169)		(0.168)			(0.173)		
SD error terms at level 3						0.086		
(SE)						(0.042)		
Observations ('person years')	1,557		1,557			1,557		
Number of persons	556		556			556		
Number of preschools						159		

Notes: Based on ten fully imputed data sets, Robust standard errors in parentheses, SD: Standard deviation, HH: household.

Sources: doi:10.5157/NEPS:SC2:8.0.1; authors' calculations.

Table A3: Additional growth curve models on vocabulary skills – extensions of Model 3e, Table 5

	Model A4		Model A5			Model A6		
Variables	**Intercept**	**Slope**	**Intercept**	**Slope**	**Slope^2**	**Intercept**	**Slope**	**Slope^2**
Intercept/Slope (Age)	−0.936	0.402	−0.888	0.453	−0.015	−0.887	0.453	−0.014
	(0.081)	(0.027)	(0.075)	(0.029)	(0.004)	(0.075)	(0.029)	(0.004)
	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.001
Education mother (Ref. CASMIN 1 a, b, c)								
CASMIN 2a, b	0.169	0.057	0.150	0.055		0.144	0.055	
	(0.072)	(0.023)	(0.071)	(0.023)		(0.071)	(0.023)	
	0.019	0.015	0.035	0.015		0.042	0.015	
CASMIN 2c_gen, d	0.251	0.060	0.239	0.058		0.235	0.058	
	(0.078)	(0.027)	(0.075)	(0.026)		(0.075)	(0.026)	
	0.001	0.024	0.001	0.024		0.002	0.024	
CASMIN 3a, b	0.339	0.069	0.286	0.070		0.278	0.070	
	(0.084)	(0.029)	(0.079)	(0.026)		(0.079)	(0.026)	
	0.000	0.016	0.000	0.008		0.000	0.008	
Education father (ref. same level as mother)								
higher than mother	0.118	−0.006						
	(0.055)	(0.019)						
	0.032	0.739						
lower than mother	0.032	−0.006						
	(0.052)	(0.017)						
	0.535	0.708						
no father (= mother lone parent)	0.102	−0.004						
	(0.077)	(0.027)						
	0.187	0.878						
Household possessions	0.090	0.009	0.098	0.009		0.097	0.009	
	(0.024)	(0.008)	(0.024)	(0.008)		(0.023)	(0.008)	
	0.000	0.250	0.000	0.249		0.000	0.251	
Reading to child	0.039	0.013	0.039	0.014		0.039	0.014	
	(0.021)	(0.007)	(0.021)	(0.007)		(0.021)	(0.007)	
	0.060	0.050	0.064	0.044		0.065	0.044	
Girl (ref. boy)	0.001	0.022	0.004	0.019		0.004	0.019	
	(0.039)	(0.013)	(0.039)	(0.013)		(0.039)	(0.013)	
	0.988	0.087	0.918	0.126		0.919	0.127	
Immigration status (no)								
At least one grandparent born abroad	−0.173	0.024	−0.178	0.025		−0.175	0.025	
	(0.054)	(0.017)	(0.054)	(0.017)		(0.054)	(0.017)	
	0.001	0.163	0.001	0.145		0.001	0.143	
Language at home (teacher report; ref. German)								
Non-German	−0.710	−0.012	−0.693	−0.014		−0.696	−0.014	
	(0.099)	(0.033)	(0.099)	(0.033)		(0.099)	(0.033)	
	0.000	0.726	0.000	0.667		0.000	0.664	

(*Continued*)

Table A3: (*Continued*)

	Model A4		Model A5			Model A6		
Variables	**Intercept**	**Slope**	**Intercept**	**Slope**	**Slope^2**	**Intercept**	**Slope**	**Slope^2**
Number of children under age 14 in HH (logarithm)	−0.139	−0.017	−0.141	−0.017		−0.137	−0.017	
	(0.051)	(0.017)	(0.050)	(0.016)		(0.050)	(0.016)	
	0.006	0.311	0.005	0.303		0.007	0.302	
Random effects								
SD random error term at level 1	0.305		0.304			0.304		
(SE)	(0.010)		(0.010)			(0.010)		
SD error terms at level 2 (person level)	0.346	0.082	0.348	0.082	fix	0.339	0.082	fix
(SE)	(0.019)	(0.010)	(0.019)	(0.010)	fix	(0.021)	(0.010)	fix
Correlation of errors at level 2	0.528		0.537			0.544		
(SE)	(0.172)		(0.172)			(0.177)		
SD error terms at level 3						0.081		
(SE)						(0.043)		
Observations ('person years')	1,557		1,557			1,557		
Number of persons	556		556			556		
Number of preschools						159		

Notes: Based on ten fully imputed data sets, Robust standard errors in parentheses, SD: Standard deviation, HH: household.

Sources: doi:10.5157/NEPS:SC2:8.0.1; authors' calculations.