Title: No Matthew effects and stable SES gaps in math and language achievement growth throughout schooling: Evidence from Germany

Abstract: The extent to which achievement gaps become wider or narrower over the course of schooling is a topic that is widely discussed, both publicly and in educational research. This study examines whether absolute achievement (in language and math skills) and social origin gaps grow throughout the school career. To investigate the achievement growth of three German student cohorts (N = 14,273) at different stages of their school career (primary school; lower secondary school; and upper secondary school), I use multilevel models to estimate the effects of prior achievement and social origin on achievement growth. The results consistently suggest a negative association between prior achievement and subsequent growth: hence, initially low-performing students have higher achievement gains than initially high-performing students. Additionally, I find that social origin gaps remain stable over time. However, when controlling for initial achievement, slightly growing socioeconomic status gaps can be observed.

1 Introduction

Mathematics skills and language skills are important predictors of school and later career success (Artelt et al., 2013; Ritchie and Bates, 2013). Inequalities in these skills can have long-lasting consequences and impact crucial school and school-to-work transitions (Linberg et al., 2019; Heckman, 2006). Differences in achievement and achievement growth have therefore been extensively examined. In the sociological literature, this topic is often addressed in relation to social inequality (von Hippel et al., 2018) or inequality-reinforcing mechanisms (DiPrete and Eirich, 2006; Stanovich, 1986; Baumert et al., 2012). Several studies show that, even before entering school, there are marked differences in achievement between children that are strongly influenced by social origin (Heckman, 2006; Skopek and Passaretta, 2021). Whether these inequalities increase or decrease over the course of schooling is therefore a central question in the sociology of education (von Hippel et al., 2018). As a result, researchers often investigate the effect of prior achievement on subsequent achievement and its interrelations with social origin (Dumont and Ready, 2020; Ready, 2013).

Two methodological approaches dominate the research to examine inequalities in achievement over time. The relative approach examines positions in a distribution (e.g. ranking in an achievement distribution) in order to draw conclusions about rank order stability between individuals and social groups over time, or about the relative distance between groups in a distribution (Skopek and Passaretta, 2021; Jerrim and Vignoles, 2013). The absolute approach examines achievement differences (e.g. gaps in proficiency) between individuals, groups, and points in time. Thus, the focus of this method is on quantifying the extent to which achievement growth over time (for example: do children improve their academic skills in school and, if so, how much?) can be detected, and how large absolute achievement gaps between individuals and groups are (Ready, 2013).

Possible mechanisms for differential achievement development depending on prior achievement and socioeconomic status (SES) are referred to as *Matthew effects*, *cumulative advantages*, *compensation effects*, or *ceiling and catch-up effects* (DiPrete and Eirich, 2006; Stanovich, 1986; McCall et al., 2006; Baumert et al., 2012; Merton, 1968). To identify such mechanisms empirically, it is useful to apply the absolute approach, which allows for quantifying growth in achievement over time (e.g. language proficiency). A direct identification of the relationship between prior achievement, SES, and achievement growth is therefore feasible.

Previous research on absolute achievement gains has yielded mixed evidence regarding the extent to which prior achievement is beneficial for achievement development (e.g. Pfost et al., 2014; Dumont and Ready, 2020; Ready, 2013). In most cases, research from the German context tends to point to zero effects or negative effects between prior achievement and achievement growth (Baumert et al., 2012; Neuendorf et al., 2020). However, some studies also indicate a positive association between prior achievement and achievement growth (Pfost et al., 2012; Murayama et al., 2013). The state of research concerning whether SES achievement gaps widen during schooling is also mixed (von Hippel et al., 2018; Dumont and Ready, 2020; for Germany, see Skopek and Passaretta, 2021). Earlier studies often report growing SES gaps (e.g. McCall et al., 2006), while more recent studies mainly report stable SES gaps (e.g. von Hippel and Hamrock, 2019). One reason for this difference could be that recent studies use more appropriate test scores with scaling based on item response theory (IRT) (von Hippel and Hamrock, 2019). Recent studies from Germany on relative achievement differences using IRT-based test scores suggest stable relative SES gaps (Skopek and Passaretta, 2021; Passaretta et al., 2022). However, for Germany, studies on the development of absolute SES differences based on nationally representative data are scarce.

Germany is renowned for its highly stratified and socially selective school system (Allmendinger, 1989) characterized by early ability tracking, which occurs at the age of 10 or 12. (Early) ability tracking in particular is often seen as a reason for widening SES and achievement inequalities (Esser and Relikowski, 2015). Therefore, in the German context, it is especially interesting to investigate the extent to which development in absolute achievement growth varies across institutional settings – primary school; (tracked) lower secondary school; and upper secondary academic education – as it is conceivable that the impact of inequality-generating mechanisms might differ over the school career (Neuendorf et al., 2020).

Several limitations can be identified in the existing research regarding absolute achievement development in Germany. First, previous studies have often focused on achievement development in early childhood, for example in kindergarten or elementary school (Neumann et al., 2014; Ditton and Krüsken, 2009; Baumert et al., 2012; Pfost et al., 2012; Schneider, 2013; Fleckenstein et al., 2019). Second, most studies examine achievement gaps for relatively short observation periods (e.g. two school years; Lehmann et al., 2001; Becker et al., 2006; Bos and Scharenberg, 2010). Third, many studies use samples from single federal states, rather than nationally representative samples (Murayama et al., 2013; Schnabel et al., 2002; Neuendorf et al., 2020).

This paper aims to contribute to, and extend, the existing body of research by examining absolute achievement gaps in the domains of mathematics and language skills, measured by IRT-based test scores, as a function of prior achievement and social background across different periods (*each spanning four years*) throughout the entire school career – elementary school (Grades 1–4, ages 6–10); lower secondary school (Grades 5–9, ages 11–14); and upper secondary school (Grades 9–12, ages 15–18) – using nationally representative large-scale data from the National Educational Panel Study (NEPS) in Germany (Blossfeld and Roßbach, 2019). I address two questions in this paper. First, do the achievement gaps between low-performing and high-performing students grow over time? Second, do social inequalities in academic achievement increase over time?

In the next sections, I briefly explain my theoretical considerations and discuss the state of research. Afterward, I present the data and the statistical methods used before reporting the results. Finally, I discuss the results in the context of the current state of research.

2 Theoretical Considerations and Previous Research

In their seminal work, DiPrete and Eirich (2006) distinguish two types of cumulative advantage processes that are frequently examined in sociology: *strict (path-dependent)* cumulative advantages and *status-dependent* cumulative advantage processes. Path-dependent cumulative advantages are processes in which prior achievement has a positive causal effect on subsequent achievement growth. In contrast, status-dependent cumulative advantages imply processes by which belonging to a certain status group influences subsequent growth. In this paper, I am interested in both strict and status-dependent cumulative advantages. Even though these mechanisms have been distinguished in theory, they are evidently interrelated, as there is a strong link between achievement and social origin.

We can distinguish three possible connections between initial achievement and subsequent growth (Dumont and Ready, 2020; Pfost et al., 2014), as illustrated in Figure 1: first, initial achievement has a positive influence on achievement growth (a, cumulative advantage); second, initial achievement and achievement growth are unrelated (b, constant development); and third, initial achievement has a negative influence on achievement growth (c, compensation effects).

Figure 1: Possible achievement growth curves by achievement quartiles at the first measurement point

— Figure 1 about here —

Source: Author's depiction.

The first relationship (a) is discussed using many different keywords, such as *cumulative advantages*, *virtuous cycles*, *success breeds success*, *Matthew effects*, or *learning begets learning* (Merton, 1968; DiPrete and Eirich, 2006; Cunha and Heckman, 2007). A cumulative advantage effect assumes that initially high-scoring students achieve stronger growth than initially low-scoring students. Concerning reading development, Stanovich (1986) postulates such a cumulative advantage when he describes the Matthew effect of reading: students who are initially better at reading improve faster than poor readers because reading promotes vocabulary knowledge, vocabulary knowledge promotes reading comprehension, and reading comprehension promotes reading skills. The gap between better-performing and weaker-performing students therefore widens over time. Since mathematical skills are also built on cumulative knowledge, such path-dependent cumulative advantages may also be expected for this domain (Neuendorf et al., 2020). In their model of skills formation, Cunha and Heckmen (2007, p. 35) also assume that, regardless of the specific domain, "skills beget skills".

In contrast, in the case of compensation effects (c), the literature assumes that the initial gap will narrow over time (Dumont and Ready, 2020; Pfost et al., 2014), be it through ceiling effects or through lack of support for very good pupils in the school system. Ceiling effects can occur, for example when students reach a performance plateau because they have reached proficiency in a domain or because certain more advanced topics are not covered in class (Baumert et al., 2012). This could lead to students who have not yet reached this plateau compensating for their lag. On the other hand, compensation effects could also occur because teachers might not provide targeted support to particularly good students. There is evidence that teachers base their instruction on an imaginary average student rather than on the actual level of the class (Archambault et al., 1993). This could result in good students benefiting less from instruction, for example because they have already mastered the material, and in previously weaker students being able to compensate for previous poor performance.

The third possible relationship – a constant rate of development (b) – emphasizes the importance of reducing inequalities before entering the school system, as inequalities remain relatively constant thereafter. It is often argued that differences between pupils usually emerge very early in life and then remain constant (Heckman, 2006). This developmental pattern could occur, for example, if all children were to reach an individual achievement plateau very early in the life-course (Baumert et al., 2012). However, an individual performance plateau established so early in the life-course (before the age of six) seems unlikely, given the empirically observed growth rates in language and mathematics skills (Bloom et al., 2008). A constant rate of development could also occur if individual cumulative advantages were consciously or unconsciously dampened, for example by the school system, by the curriculum, or by the teachers.

From a variety of studies, we know that a higher SES is associated with higher academic achievement. This association can already be observed very early in life (Heckman, 2006; Kulic et al., 2019). The literature proposes several mechanisms to explain why there might be a status-dependent cumulative advantage for high SES children over low SES children: higher parental involvement (Marks et al., 2006); better homework assistance (Pfeffer, 2008); class-specific socialization processes (Nash, 2003); higher economic resources and human capital (Bernardi, 2014); and efforts by high SES families to avoid downward mobility (Lucas, 2001). International studies have revealed that the achievement gap between social groups either widens over time (McCall et al., 2006; Dumont and Ready, 2020) or remains largely constant (Skopek and Passaretta, 2021; von Hippel et al., 2018). The differences in research findings could be caused by using different test scales (von Hippel and Hamrock, 2019). They could also be caused by researchers choosing different statistical models to answer different questions (e.g. conditioning vs non-conditioning on prior achievement in achievement growth studies) (Ready, 2013; Kelly and Ye, 2017). By conditioning on prior achievement, we are able to examine the extent to which students from different backgrounds develop given the same initial achievement. However, in analyses without conditioning on initial achievement, we can ask how the differences between the mean high SES and low SES children develop over time. In addition to the different epistemological aims of these approaches, the following methodological differences must also be noted. Taking prior achievement into account may lead to an overestimation of SES effects, for example due to measurement error in achievement measures (Passaretta et al., 2020; Jerrim and Vignoles, 2013). However, not taking this association into account may lead to underestimating SES effects (Baumert et al., 2012; Ditton and Krüsken, 2009).

If path-dependent cumulative advantages are present, achievement gaps by SES at the beginning of schooling could also lead to widening SES gaps even if no status-dependent cumulative advantages influence this process. On the other hand, path-dependent compensation effects might mask status-dependent cumulative advantages (e.g. if constant SES gaps are observed: Baumert et al., 2012; Ditton and Krüsken, 2009). To separate such path-dependent processes from status-dependent processes, both processes need to be analyzed simultaneously (Baumert et al., 2012).

Theoretically, it is also conceivable that SES gaps might be reduced over time through targeted interventions such as early childhood education, free childcare, reduced educational costs for low SES families, education vouchers, or special tutoring programs (Kelly and Ye, 2017). However, although it is theoretically possible to reduce SES gaps, it must be noted that (unlike SES gaps that are either constant or growing), there is hardly any empirical evidence of this happening.

Table 1 shows the state of research on the influence of prior achievement and/or SES on achievement growth in Germany. Studies have only been included if they longitudinally examined differences in academic achievement in language skills (vocabulary or reading) or mathematics skills in connection with prior achievement and/or SES. The state of research on path-dependent and status-dependent cumulative advantages is not conclusive for Germany. However, the findings indicate small compensation effects rather than cumulative advantages. No study depicts decreasing status-dependent SES effects. Rather, the evidence points to constant or slightly widening SES gaps. From the state of research, it appears that these results are very similar for the domains of mathematics and language skills.

— Table 1 about here –

Based on the state of research and the theoretical considerations presented above, I hypothesize the following relationship between prior achievement and achievement growth.

Hypothesis 1: Initially low-performing students experience higher achievement growth than initially high-performing students (compensation effects).

As indicated above, status-dependent cumulative advantages could be masked by path-dependent cumulative effects. A negative relationship between initial achievement and subsequent growth could counteract any positive SES effects. Which of the two cumulative advantage effects is stronger (e.g. whether we observe constant or increasing SES effects) must be tested empirically. In line with the current state of research, after controlling for the path-dependent process (negative initial achievement growth association), it can be assumed that positive SES effects will occur (Baumert et al., 2012; Ditton and Krüsken, 2009; Dumont and Ready, 2020).

Hypothesis 2: High SES students experience higher achievement growth than low SES students (controlling for prior achievement).

In the German context[1], with its early ability tracking and highly stratified secondary school tracks, the effects described above are expected to vary across different institutional settings – primary school; (tracked) lower secondary school; and upper secondary academic education. As stated above, I assume a negative relationship between prior achievement and achievement growth. This relationship could vary between different institutional settings, as previous research points to decreasing learning rates over the school career (Bloom et al., 2008). The older students get, the closer they get to their learning plateau. Accordingly, compensation effects caused by learning plateaus are more likely to emerge in secondary school than in primary school. However, ability tracking in secondary school – which results in students with higher initial achievement being more likely to attend school tracks with more challenging curricula (Esser and Relikowski,

2015) – could have a counteracting effect, widening achievement gaps by prior achievement and reducing the assumed compensation effects for these students. Although the expected compensation effects may be reduced by ability tracking, I still expect them to be higher than in primary school because the children are older. Since students in academic upper secondary schools are a performance-selective group of the student body, who are aiming for the highest school leaving qualification and who are already relatively old (aged 15–16), I assume that compensation effects caused by learning plateaus should hardly appear during this phase of schooling. Following the argumentation of learning plateaus as a reason for compensation effects, I expect compensation effects to be most pronounced in lower secondary school.

Hypothesis 3: Compensation effects should be most evident in lower secondary school.

As noted above, it can be assumed that achievement growth rates decrease over the course of schooling (Bloom et al., 2008). Because high SES students enter school with an achievement advantage, these students should reach learning plateaus earlier than low SES students[2]. Therefore, the positive association between SES and achievement growth – controlled for path-dependent cumulative advantage processes – should decrease across institutional settings. This decrease could be counteracted by the selective student group in upper secondary grammar schools and by higher growth rates for students following the most demanding secondary school tracks, in which high SES students are overrepresented (Esser and Relikowski, 2015; Murayama et al., 2013). For Germany, the state of research on absolute SES gaps after tracking is scarce. However, based on the state of research on relative SES gaps, I assume that the positive SES effects should be most evident in primary school (Skopek and Passaretta, 2021).

Hypothesis 4: The positive association between SES and achievement growth should be most pronounced in primary school (controlling for prior achievement).

3 Data, Measures, and Statistical Procedure

4.1 Data and Sample

In this paper, I use data from three starting cohorts (SCs) from Germany's NEPS: SC2, SC3, and SC4[3]. The NEPS is a multi-cohort study that monitors respondents throughout their educational career and surveys them at regular annual intervals. In addition to the students, the parents are also interviewed. For this study, I use data from cohorts that were interviewed for the first time in kindergarten (SC2), at the beginning of lower secondary school (SC3), and at the end of lower secondary school (SC4). The target population for these surveys consisted of all pupils attending Grade 1 during the school year 2012/13 (SC2) and all pupils attending either Grade 5 (SC3) or Grade 9 (SC4) at secondary school in the school year 2010/11. The samples were drawn from the target population using stratified sampling (for further information on the sampling procedure, see Aßmann et al., 2011). For SC2, I use data from 2012–16, the years the children attended primary school (Grades 1–4). For SC3, I use data from 2010–15, the years the children attended lower secondary school (Grades 5–9). For SC4, I use data from 2010–14, the years the children attended upper secondary school (Grades 9–12). Since the academic track is the only one that goes up to Grade 12, and since the children who did not pursue the academic track were not tested after leaving school, I have only included the children from grammar schools (*Gymnasium*) in the sample of this final cohort.

For my analytical sample, I focus on children with no special needs who participated in at least the first or last wave of the survey, and who remained at the same school during the observation period. This sample restriction is imposed because children who changed schools did not have their achievement individually tested (Aßmann et al., 2011). Therefore, pupils from the federal states of *Brandenburg*, *Berlin*, and *Mecklenburg-Vorpommern* were excluded from the SC3 sample because, in those states, students transition to lower secondary school after Grade 6 and are thus considered school changers in NEPS from Grade 7 onward, so competency scores are not available for them. The resulting analytical sample includes 14,273 students in total for mathematics ($N_{\text{primary school}}$ = 6,866; $N_{\text{lower secondary school}}$ = 3,461; $N_{\text{upper secondary school}}$ = 3,946) and

14,123 students in total for language skills ($N_{primary\ school} = 6{,}865$; $N_{lower\ secondary\ school} = 3{,}350$; $N_{upper\ secondary\ school} = 3{,}908$).

4.2 Measures

I use the students' scores in mathematics and language skills as outcomes in this paper. Individual achievement growth was constructed by subtracting the score achieved in the previous achievement test (SC2: Grade 1; SC3: Grade 5; SC4: Grade 9) from the score achieved in the following test (SC2: Grade 4 [for language skills Grade 3]; SC3: Grade 9; SC4: Grade 12).

The achievement tests in NEPS are designed to observe changes in achievement from a life-course perspective. The mathematics tests measure content (e.g. quantity; change and relationships; space; and shape) and cognitive components (e.g. modeling; mathematical problem solving) (Neumann et al., 2013). The reading skill tests measure cognitive requirements (e.g. finding information; drawing conclusions; reflecting on the text) for different text types and functions (e.g. literature; instructions; advertising; information; and commentary) (Gehrer et al., 2013). Since weighted likelihood estimates (WLE) scores are only available for the reading achievement of SC2 from Grade 4 on, I use the WLE scores of receptive vocabulary achievement from Grade 1 and Grade 3 for SC2. NEPS uses a test based on picture selection tasks (a modified version of the Peabody Picture Vocabulary Test) to measure receptive vocabulary (Fischer and Durda, 2020), which "comprises all words a person recognizes and comprehends when heard" (Berendes et al., 2013, p. 36).

The achievement measurements in NEPS are either based on an anchor group or on an anchor item design for successive measurement points (Pohl and Carstensen, 2012; Fischer et al., 2016). I use WLE scores provided by NEPS, which are suitable for longitudinal comparisons (Fischer et al., 2016). The WLE estimates are based on IRT models, and their reliability can be considered as good (see Table B1 in the Supporting Online Material [SOM]). Von Hippel and Hamrock (2019) show that test scores based on IRT scaling models are preferable to test scores based on other scales (e.g. number-right or Thurstone scales) when examining achievement gaps. These achievement scores are comparable inter-individually and intra-individually within a student cohort, but not between cohorts. Further details on the achievement tests are provided in Chapter B of the SOM.

To construct the social origin indicator, I use the highest International Socio-Economic Index of Occupational Status (ISEI) of the parents (Ganzeboom and Treiman, 1996)[4]. This scale ranges from 10 to 90. I have divided it into three equally broad groups: low SES (10–36); middle SES (37–63); and high SES (64–90). Achievement studies often subdivide social background into three groups (e.g. Skopek and Passaretta, 2021)[5]. Information about the parents' occupations was taken from the survey filled in by the parents. If no such information appeared in the parents' interview, information from the children's interview was used. Students were attributed a migration background if they belonged to the 1st–3.5th migrant generation in Germany. I include this variable in my models because minority students have lower achievement gains than majority students (McCall et al., 2006; Relikowski et al., 2015). Furthermore, I include a dummy variable indicating student gender in my analyses because previous research has indicated that boys make higher gains in mathematics and that girls make higher gains in reading (Ehrtmann and Wolter, 2018). As additional controls, I include age at first measurement and the distance between first and last measurement in months in the models.

4.3 Analytical Strategy

There are different statistical approaches to estimate the association between prior achievement and subsequent achievement. On the one hand, the status attainment model (Equation 1) regresses the test score of measurement point t on that of measurement point t-1; on the other, the regressor variable model (Equation 2) regresses the subsequent growth between two measurement points (G) on the prior achievement. As various authors have shown, the two models are "algebraically equivalent" (Kelly and Ye, 2017 p. 357). Only the reference point of the effect of γ_1 differs. In the status attainment model, any non-zero association between prior achievement and subsequent growth is represented by a deviation of 1, while in the regressor variable models (Equation 2) this is represented by a deviation of 0 (Allison, 1990; Kelly and Ye, 2017). This relationship between the two models becomes clear when $Y_{t\text{-}1}$ is subtracted from Model 1 on both sides (Equation 3)[6],

$$Y_t = \gamma_0 + \gamma_1 Y_{t-1} + \gamma_X \mathrm{X} + r \qquad (1)$$

$$\mathrm{G} = Y_t - Y_{t-1} = \gamma_0 + \gamma_1 Y_{t-1} + \gamma_X \mathrm{X} + r \qquad (2)$$

$$Y_t - Y_{t-1} = \gamma_0 + \gamma_1 Y_{t-1} - Y_{t-1} + \gamma_X \mathrm{X} + r = \gamma_0 + (\gamma_1 - 1) Y_{t-1} + \gamma_X \mathrm{X} + r \qquad (3)$$

where

γ are the regression coefficients;

Y_t is the achievement score at the last measurement;

Y_{t-1} is the achievement score at the first measurement;

X is a vector for all the individual-level model variables, including SES; and

r is the error term.

Since the regressor variable model (Equation 2) provides a direct estimator of the effect of prior achievement on growth, I estimate this model to answer the first research question about the relationship between prior achievement and subsequent growth. The regressor variable model examines achievement growth as a linear function of a student's initial achievement score. If the parameter γ_1 is above 0, this can be interpreted as a cumulative advantage; if the parameter is below 0, this is read as a compensation effect (Baumert et al., 2012; Pfost et al., 2014; DiPrete and Eirich, 2006)[7].

Because of the nested data structure – students (i) are nested in schools (j) – I estimate the regressor variable model (see Equation 2) as a multilevel mixed model (see Equation 4)[8]. I expect the achievement to differ between schools, which is why I include random intercepts at the school level (u0j). Furthermore, because curricula and achievement differences vary between federal states (Naumann et al., 2010), I include federal states dummy variables in my model (Zij). For data protection reasons, the effects of the federal states are not subsequently displayed.

To answer the second research question concerning the relationship between achievement and social inequalities, different approaches are used in the literature (Dumont and Ready, 2020;

Ready, 2013). One approach considers prior achievement as an explanatory variable in the analyses (regressor variable model, as in Equation 4) and attempts to assess the extent to which students from different social backgrounds develop given the same baseline achievement. Another approach does not include prior achievement as an explanatory variable in the analyses (as in Equation 5) and attempts to assess the extent to which students from different social backgrounds differ in their achievement growth. This model is referred to in the literature as a change-score model (Allison, 1990). These approaches highlight different comparisons (conditional [Equation 4] and unconditional [Equation 5] on prior achievement) and thus attempt to answer different questions (Dumont and Ready, 2020; Ready, 2013). In order to obtain a comprehensive idea of the relationship between SES and inequalities in achievement growth, both approaches are relevant (Baumert et al., 2012). Therefore, both regressor variable models and change-score models for the SES effects are calculated in the following.

$$G_{ij} = Y_{tij} - Y_{t-1ij} = \gamma_{00} + u_{0j} + \gamma_{10} * Y_{t-1ij} + \gamma_{X0} * X_{ij} + \gamma_{Z0} * Z_{ij} + r_{ij} \quad (4)$$

$$G_{ij} = Y_{tij} - Y_{t-1ij} = \gamma_{00} + u_{0j} + \gamma_{X0} * X_{ij} + \gamma_{Z0} * Z_{ij} + r_{ij} \quad (5)$$

All missing data were replaced by using multiple imputation with chained equations (applying predictive mean matching using the 10 nearest neighbors), producing 50 complete data sets[9]. The descriptive summary statistics of the analytical samples before and after the imputation are displayed in Tables C1–C4 in Chapter C in the SOM. I used Stata 17 for all data preparations, for multiple imputation, and for the analysis (Jann, 2004; Jann, 2007). The code can be found in the Additional Online Material. Prior achievement, age at first measurement, and distance between measurements were grand mean-centered within each of the imputed data sets.

4 Bivariate Results

For all three cohorts and both domains (mathematics and language skills), we can observe a negative correlation between prior achievement and subsequent growth (see Table 2). This points to a compensation effect in the development of achievement[10].

— Table 2 about here —

Figure 2 shows the average achievement by social origin of the three student cohorts at the different measurement points. Two things become visible in this graph. First, in each cohort, achievement is stratified according to social background: the higher the SES, the higher the achievement. Second, the growth rates are positive and quite similar for all status groups. For a better interpretation of the achievement gaps between the SES groups, I calculate effect sizes (Cohen's delta) between high SES and low SES students for various points in time (see Table 3). In both domains, the effect sizes in primary and lower secondary school range from .82 to 1.03. The group differences are substantial and hardly change over time. In upper secondary school, the effect sizes are smaller (between .22 and .34), which is certainly due to the selective student body, but also highly stable over time. These results suggest persistent inequalities by SES.

Figure 2: Achievement development over time by social origin, student cohort, and domain: a) mathematics, b) language skills

— Figure 2_1 about here —

— Figure 2_2 about here —

Source: NEPS SC2(9.0.0), SC3(11.0.1), SC4(12.0.0); based on 50 multiple-imputed data sets, author's calculations.

— Table 3 about here —

5 Multilevel Results

Figure 3 shows the regression coefficients (change-score and regressor variable models) of prior achievement and social origin on achievement growth for mathematics and language skill achievement (see Chapter C.2 in the SOM for the multilevel regression tables and the discussion of the effects of the control variables).

The multilevel models confirm the negative association between prior achievement and subsequent growth. In all three cohorts and for both domains, the parameter for this relationship is negative and statistically significant. The higher the prior achievement, the lower the subsequent growth. This result is in line with Hypothesis 1: there are compensation effects in achievement development. However, Hypothesis 3 is not supported by the results. Apart from the language skills in SC2, the compensation effects seem very similar across institutional settings.

The results of the change-score models show small significant SES effects in primary school and no substantial positive SES effects in secondary school. However, the results of the regressor variable models show significant positive effects of SES on achievement growth when controlling for prior achievement in all three institutional settings. These results are in line with Hypothesis 2.

In line with Hypothesis 4, the SES gaps in both domains – both when prior achievement is controlled for, and when it is not – diverge most in terms of magnitude in primary school. However, the SES gaps also increase in lower and upper secondary school once prior achievement is controlled for. Overall, the SES effects can be classified as rather small.

Figure 3: Prior achievement and SES coefficients on achievement growth (with 95% CI), by domain and student cohort

— Figure 3 about here —

Note: Change-score model (based on Equation 5; unconditional on prior achievement). Regressor variable model (based on Equation 4; conditional on prior achievement). Full estimation results displayed in Tables C5 and C6 in the SOM.

Source: NEPS SC2(9.0.0), SC3(11.0.1), SC4(12.0.0); based on 50 multiple-imputed data sets, author's calculations.

6 Robustness Checks

I conducted a series of robustness checks, which are described and presented in detail in Chapter D in the SOM. First, I used a sample consisting only of students for whom the dependent variable did not need to be imputed (WoDV). Second, I estimated models in which an intelligence test score was additionally included as a control variable (+Reasoning). Third, I used a sample in which only the central 60% of the first achievement measure was included to mitigate possible bias due to ceiling effects or sensitivity of the competence assessments in the extreme ranges (Central 60).[11] The results were virtually identical to those reported in the main analysis.

Regression to the mean effects induced by measurement error pose a threat to my conclusions of prior achievement and SES effects. Conditioning on the first achievement measure containing measurement error (very good students can only get worse in subsequent measures, and very bad students can only get better) might lead to underestimating the effects of prior achievement and overestimating the SES effects. Thus, both the reported compensation effects and the positive SES effects could be the result of statistical artifacts. I ran several models to mitigate the possible effects of measurement error. First, I used the achievement rank instead of the absolute achievement of the first measurement time point (Achievement rank). Second, I used the achievement rank as an instrument for achievement in a two-stage model (IV). Third, I regressed the mean achievement of the first two measurements on the achievement growth between the second and third measurement (only for cohorts with three measurement time points [Mean achievement]). Fourth, I calculated other model specifications, such as fixed effect growth curve models with an interaction between time and SES and time and prior achievement (only for cohorts with three measurement time points [FE]). Figure 4 summarizes the results of these robustness checks. None of these models led to substantially different conclusions from those presented above. However, as expected, the effects of SES and prior achievement of the robustness check models tended to be closer to zero than in the models presented above.

All results for SC3 presented in the paper were based on models that did not consider the school track. However, even when the school track was included in the models for SC3 (Murayama et al., 2013), the results did not change substantially (see, for example, Tables C5 and C6 in the SOM). It is interesting to note that there were hardly any substantial differences between the tracks in the models that do not take initial achievement into account. However, once initial achievement was controlled for, students following the academic track experienced considerably higher achievement growth than students on the other tracks.

Figure 4: Robustness checks: prior achievement and SES coefficients (high SES vs low SES) on achievement growth (with 95% CI), by domain, student cohort, and estimation model

— Figure 4_1 about here —

— Figure 4_2 about here —

Note: Regressor variable approach models only. Full estimation results displayed in Chapter D in the SOM. Mean achievement models and fixed effect models estimated only for cohorts with three measurement time points.

Source: NEPS SC2(9.0.0), SC3(11.0.1), SC4(12.0.0); based on 50 multiple-imputed data sets, author's calculations.

7 Discussion and Conclusion

The aim of this study has been to examine the relationship between prior achievement, SES differences, and achievement growth within the German school system. To this end, I have examined the absolute achievement development in language skills and mathematics of three student cohorts in different institutional contexts (primary school; lower secondary school; and upper secondary school), using large-scale data from different NEPS cohorts. Following previous research, I have used regressor variable (conditioning on prior achievement) and change-score (no conditioning on prior achievement) models to identify cumulative advantage effects.

The first research question sought to determine whether path-dependent cumulative advantages in achievement development can be detected and, if so, the extent to which they vary across different institutional contexts. I have been able to show a negative relationship between prior achievement and subsequent growth in each of the three student cohorts for both domains. Compensation effects are evident in the German school system. Initially lower-scoring students make higher gains than initially higher-scoring students. These results support Hypothesis 1.

Nevertheless, it is worth noting that there are already substantial differences in achievement at the beginning of a child's school career (Heckman, 2006; Linberg et al., 2019). Although the gap reduces slightly over the course of schooling, substantial differences remain. Hypothesis 3, which assumed that compensation effects mainly appear in lower secondary school, could not be confirmed. The compensation effects were relatively similar in all institutional settings. The only exception was the lower compensation effect in language skills in primary school. These smaller effects could be because receptive vocabulary was examined for this period rather than reading skills, which were examined during the other periods. Previous studies have demonstrated compensation effects in absolute achievement gaps for shorter study periods of one to two school years during the school career or in selected federal states (see Table 1). This study has contributed to the state of research by demonstrating compensation effects using nationally representative data for each four-year study period in each institutional context in the domains of language skills and mathematics skills, controlling for possible effect bias due to measurement error in test scores.

There might be different reasons for these compensation effects. On the one hand, learning plateaus could lead to previously good students performing less strongly (Baumert et al., 2012). The compensation effects could also be caused by teachers supporting lower-performing students more than higher-performing students, or aligning their instruction with an imaginary average student (Archambault et al., 1993), thus not supporting high-performing students to the fullest extent possible. The aforementioned reasons could not be examined in the present study, so further research is needed to investigate why these compensation effects occur.

This article's second research question pertained to status-dependent cumulative advantages over the course of schooling. Bivariate results indicated that the SES gap remained largely stable during the period under review. The results of the change-score models supported the bivariate finding that the SES gaps remain constant over time. As soon as I considered the negative association between prior achievement and subsequent growth (regressor variable models), small positive SES effects emerged. These results support Hypothesis 2. Furthermore, the results of the regressor variable models support Hypothesis 4: the strength of SES effects is greatest in primary school and decreases across institutional settings (once prior achievement has been controlled for). Positive SES effects under the control of prior achievement are not equivalent to status-dependent cumulative advantages. If the gap between the different status groups were to widen, we would be able to see this in simple descriptive graphs, but this does not happen.

These results help understand why constant SES gaps can be observed. If low SES and high SES students had the same level of prior achievement, the SES gaps would grow. However, since prior achievement is socially stratified, high SES students are more affected by the compensation effects of prior achievement. Thus, the SES gaps do not grow, but they do not decrease either, since high SES students can partially compensate for the negative effect of prior achievement. The results of the regressor variable models are in line with previous research, which has found positive SES effects under the control of prior achievement using data from selected federal states (e.g. Ditton and Krüsken, 2009; Baumert et al., 2012). The results of this study nevertheless go beyond previous research in demonstrating these effects during a longer study period for nationally representative student cohorts. The results from the change-score models – that absolute SES gaps remain largely constant or only increase slightly in primary school – support prior research on relative SES gaps, which has shown that much SES achievement inequality arises before school enrollment and remains constant thereafter (for the United States, see von Hippel et al., 2018; for Germany, see Skopek and Passaretta, 2021; for the United Kingdom, see Jerrim and Vignoles, 2013).

An answer to the question of why we observe growing SES gaps once prior achievement has been taken into account is beyond the scope of this paper and should be addressed in future research. Referring to previous studies, high SES parents may be more involved in their children's school activities; they may be more supportive of their children's homework; or they may use more economic resources to buy more learning materials for their children and provide a more engaging learning environment (Marks et al., 2006; Pfeffer, 2008; Bernardi, 2014).

As the German school system is characterized by high selectivity and early ability tracking, we might question how far the reported relationships between prior achievement, SES, and achievement growth can be generalized beyond the German context. The reported effects are in many cases similar across institutional contexts (primary school, lower secondary school, and upper secondary school). Accordingly, the differences between the institutional settings are smaller than theoretically expected, which might indicate that these results can also be generalized beyond Germany. This interpretation is also supported by the fact that the reported negative prior achievement growth association, the constant SES gaps (unconditional on prior achievement), and the growing SES gaps (conditional on prior achievement) are in line with international study results (e.g. Dumont and Ready, 2020; McCall et al., 2006; Ready, 2013; Jerrim and Vignoles, 2013). Nevertheless, statements about international generalizability based on one-country studies are difficult, and a conclusive judgment on this question requires further internationally comparative research.

Although I have computed a series of robustness checks in support of my conclusions, I cannot rule out bias due to measurement error. Accordingly, my results should be interpreted as overestimating the reported SES effects and underestimating the effects of prior achievement (Passaretta et al., 2022). However, the results of the robustness tests suggest that the bias should be relatively small, so I assume my substantive conclusions are robust.

A limitation of this study is that I only have achievement data covering a period of four years for individual students. While this is a longer period than has been dealt with in many previous studies, it also means that, while I can study a specific period of schooling for each cohort, I cannot investigate the entire school career of individual children. However, this limitation should be resolved in a few years when the children of SC2 leave the school system.

In summary, the gap in language and mathematics achievement between initially high-scoring and low-scoring students does not grow in the German school system. Low-achieving students are able to narrow the initial gap in achievement, but substantial differences remain. Furthermore, there are no status-dependent cumulative advantages for high SES students compared to low SES students, and the achievement gap does not widen based on social background. Nevertheless, inequality processes by SES are evident: given the same initial performance, high SES students show greater achievement gains than low SES students.

Notes

1. For a brief overview of the German school system, see Chapter A in the Supporting Online Material (SOM).

2. It could be argued that there are different SES learning plateaus as a result of differential learning environments. To test for such a relationship, models with interactions between prior achievement and SES were run in addition to those shown below. As in the study by Passaretta et al. (2022), which also uses NEPS data, the interaction effects are not statistically significant, and the main effects are hardly affected by the inclusion of these interactions (see Figure C3 in the SOM).

3. This paper uses data from the NEPS (see Blossfeld and Roßbach, 2019). The NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi, Germany) in cooperation with a nationwide network (NEPS Network, 2020; 2021; 2021a).

4. In studies of achievement differences by SES, various indicators are used to measure SES (ISEI; social class; parental education; income; free school lunch eligibility). Since I also rely on the children's information about their parents as not many parents participated in the survey, I decided to use the ISEI as an indicator instead of education because, in some cohorts, the children were not surveyed about their parents' education until several years after the last competency measurement I used, and some cohorts were assessed using a different scale than for their parents. Analyses using parental education as the SES indicator are hardly different from those using the ISEI (see Figure C4 in the SOM).

5. Although such categorization has been common in previous studies, the choice of the number of categories always remains arbitrary. However, analyses using the continuous ISEI (z-standardized) show very similar results to those using the categorical variable (see Figure C5 in the SOM).

6. In this model, a linear relationship between prior achievement and subsequent growth is assumed. However, it is also conceivable that this relationship is not linear and that it deviates from the linear relationship, especially at the edges of the prior achievement distribution. Additional analyses using a quartic relationship (see Figures C6 and C7 in the SOM) show no clear deviation from the linear relationship. In the peripheral areas – where, however, there are also few data points – there are minor deviations from the linear trend such that the linear trend slightly underestimates the relationship in these areas. Thus, the coefficients presented below are considered conservative estimates.

7. To study growing inequality between groups, one could also examine whether the overall variance within the population increases over time. I cannot follow this strategy because the variance of the WLE estimates was kept constant by design at all measurement points.

8. In addition to the mixed models that predominate in sociology, other statistical models can be used to account for the clustered data structure (McNeish et al., 2017), such as population average models computed using generalized estimating equations (GEE). Figure C8 in the SOM shows the results of the mixed models and of the GEE models. The choice of statistical approach has no effect on the estimation results.

9. Achievement growth was imputed by applying the just-another-variable approach. I included the following auxiliary variables in the imputation models alongside the model variables: the sampling strata; student weight; school weight; mathematics and reading achievement scores from different grades; the reasoning score; outside school reading time; parental education; school composition

measures (share of minority students, mean achievement, share of high SES students); and (for SC3 and SC4) the grades in mathematics and in German and the idealistic job aspirations (ISEI).

10. A graphical illustration of the distribution of the two variables and their interrelation is shown in Figure C1 (multiple-imputed data) and in Figure C2 (complete cases) in the SOM.

11. For a discussion of possible ceiling effects in the test scores used, see Chapter D.1 in the SOM.

References

Allison, P. D. (1990). Change scores as dependent variables in regression analysis. *Sociological Methodology*, 20, 93–114.

Allmendinger, J. (1989). Educational systems and labor market outcomes. *European Sociological Review*, 5, 231–250.

Archambault Jr, F. X., Westberg, K. L., Brown, S. W., Hallmark, B. W., Zhang, W., Emmons, C. L. (1993). Classroom practices used with gifted third and fourth grade students. *Journal for the Education of the Gifted*, 16, 103–119.

Artelt, C., Weinert, S., Carstensen, C. H. (2013). Assessing competencies across the lifespan within the German National Educational Panel Study (NEPS) – Editorial. *Journal for Educational Research Online*, 5, 5–14.

Aßmann, C., Steinhauer, H. W., Kiesl, H., Koch, S., Schönberger, B., Müller-Kuller, A., Rohwer, G., Rässler, S., Blossfeld, H. (2011). Sampling designs of the National Educational Panel Study: Challenges and solutions. *Zeitschrift für Erziehungswissenschaft*, 14, 51–65.

Baumert, J., Nagy, G., 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, 1347–1367.

Becker, M., Lüdtke, O., Trautwein, U., Baumert, J. (2006). Leistungszuwachs in Mathematik. Evidenz für einen Schereneffekt im mehrgliedrigen Schulsystem? *Zeitschrift für Pädagogische Psychologie*, 20, 233–242.

Berendes, K., Weinert, S., Zimmermann, S., Artelt, C. (2013). Assessing language indicators across the lifespan within the German National Educational Panel Study (NEPS). *Journal for Educational Research Online*, 5, 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, 74–88.

Bloom, H. S., Hill, C. J., Black, A. R., Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, 1, 289–328.

Blossfeld, H.-P., Roßbach, H.-G. (Eds.) (2019). Education as a lifelong process: The German National Educational Panel Study (NEPS). *Edition ZfE* (2nd ed.). Springer VS.

Bos, W., Scharenberg, K. (2010). Lernentwicklung in leistungshomogenen und -heterogenen Schulklassen. In Bos, W., Klieme, E., Köller, O. (Eds.), *Schulische Lerngelegenheiten und Kompetenzentwicklung. Festschrift für Jürgen Baumert.* Münster: Waxmann, pp. 173–194.

Cunha, F., Heckman, J. (2007). The technology of skill formation. *American Economic Review,* 97, 31–47.

DiPrete, T. A., Eirich, G. M. (2006). Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments. *Annual Review of Sociology*, 32, 271–297.

Ditton, H., Krüsken, J. (2009). Denn wer hat, dem wird gegeben werden? Eine Längsschnittstudie zur Entwicklung schulischer Leistungen und den Effekten der sozialen Herkunft in der Grundschulzeit. *Journal for Educational Research Online*, 1(1), 33–61.

Dumont, H., Ready, D. D. (2020). Do schools reduce or exacerbate inequality? How the associations between student achievement and achievement growth influence our understanding of the role of schooling. *American Educational Research Journal*, 57, 728–774.

Ehrtmann, L., Wolter, I. (2018). The impact of students' gender-role orientation on competence development in mathematics and reading in secondary school. *Learning and Individual Differences*, 61, 256–264.

Esser, H., Relikowski, I. (2015). Is ability tracking (really) responsible for educational inequalities in achievement? A comparison between the country states Bavaria and Hesse in Germany. *IZA Discussion Paper*.

Fischer, L., Durda, T. (2020). NEPS technical report for receptive vocabulary: Scaling results of Starting Cohort 2 for Kindergarten (Wave 1), Grade 1 (Wave 3) and Grade 3 (Wave 5). *NEPS Survey Paper No. 65*.

Fischer, L., Rohm, T., Gnambs, T., Carstensen, C. H. (2016). Linking the data of the competence tests. *NEPS Survey Paper No 1*.

Fleckenstein, J., Gebauer, S. K., Möller, J. (2019). Promoting mathematics achievement in one-way immersion: Performance development over four years of elementary school. *Contemporary Educational Psychology*, 56, 228–235.

Freund, M. J., Wolter, I., Lockl, K., Gnambs, T. (2021). Determinants of profiles of competence development in mathematics and reading in upper secondary education in Germany. *PloS one*, 16(10), e0258152.

Ganzeboom, H. B., Treiman, D. J. (1996). Internationally comparable measures of occupational status for the 1988 International Standard Classification of Occupations. *Social Science Research*, 25, 201–239.

Gehrer, K., Zimmermann, S., Artelt, C., Weinert, S. (2013). NEPS framework for assessing reading competence and results from an adult pilot study. *Journal for Educational Research Online* 5, 50–79.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312, 1900–1902.

Jann, B. (2004). *ESTOUT: Stata Module to Make Regression Tables*. Statistical Software Components, Boston College Department of Economics.

Jann, B. (2007). *FRE: Stata Module to Display One-Way Frequency Table*. Statistical Software Components, Boston College Department of Economics.

Jerrim, J., 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*, 176, 887–906.

Kelly, S., Ye, F. (2017). Accounting for the relationship between initial status and growth in regression models. *The Journal of Experimental Education*, 85, 353–375.

Kulic, N., Skopek, J., Triventi, M., Blossfeld, H.-P. (2019). Social background and children's cognitive skills: The role of early childhood education and care in a cross-national perspective. *Annual Review of Sociology*, 45, 557–579.

Lehmann, R. H., Husfeldt, V., Peek, R. (2001). Lernstände und Lernentwicklungen im Fach Mathematik: Ergebnisse der Hamburger Untersuchung (LAU) in den Jahrgangsstufen 5 und 6. In Kaiser, G, Knoche, N., Lindt, D., Zillmer, M. (Eds.), *Leistungsvergleiche im Mathematikunterricht-ein Überblick über aktuelle nationale Studien.* Hildesheim: Verlag Franzbecker, pp. 29–50.

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

Lucas, S. R. (2001). Effectively maintained inequality: Education transitions, track mobility, and social background effects. *American Journal of Sociology*, 106, 1642–1690.

Marks, G. N., Cresswell, J., Ainley, J. (2006). Explaining socioeconomic inequalities in student achievement: The role of home and school factors. *Educational Research and Evaluation*, 12, 105–128.

McCall, M. S., Hauser, C., Cronin, J., Kingsbury, G. G., Houser, R. (2006). *Achievement Gaps: An Examination of Differences in Student Achievement and Growth*. Portland, OR: ERIC.

McNeish, D., Stapleton, L.M., Silverman, R.D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*, 22, 114–140.

Merton, R. K. (1968). The Matthew effect in science. The reward and communication systems of science are considered. *Science*, 159, 56–63.

Murayama, K., Pekrun, R., Lichtenfeld, S., Vom Hofe, R. (2013). Predicting long-term growth in students' mathematics achievement: The unique contributions of motivation and cognitive strategies. *Child Development,* 84, 1475–1490.

Nash, R. (2003). Inequality/difference in education: Is a real explanation of primary and secondary effects possible? *The British Journal of Sociology*, 54, 433–451.

Naumann, J., Artelt, C., Schneider, W., Stanat, P. (2010). Lesekompetenz von PISA 2000 bis PISA 2009. In Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M., Schneider, W., Stanat, P. (Eds.), *PISA 2009. Bilanz nach einem Jahrzehnt*. Münster: Waxmann Verlag, pp. 23–71.

NEPS Network. (2020). National Educational Panel Study, Scientific Use File of Starting Cohort Kindergarten. Leibniz Institute for Educational Trajectories (LIfBi), Bamberg. https://doi.org/10.5157/NEPS:SC2:9.0.0

NEPS Network. (2021). *National Educational Panel Study, Scientific Use File of Starting Cohort Grade 5*. Leibniz Institute for Educational Trajectories (LIfBi), Bamberg. https://doi.org/10.5157/NEPS:SC3:11.0.1

NEPS Network. (2021a). National Educational Panel Study, Scientific Use File of Starting Cohort Grade 9. Leibniz Institute for Educational Trajectories (LIfBi), Bamberg. https://doi.org/10.5157/NEPS:SC4:12.0.0

Neuendorf, C., Jansen, M., Kuhl, P. (2020). Competence development of high achievers within the highest track in German secondary school: Evidence for Matthew effects or compensation? *Learning and Individual Differences*, 77, 101816.

Neumann, I., Duchhardt, C., Grüßing, M., Heinze, A., Knopp, E., Ehmke, T. (2013). Modeling and assessing mathematical competence over the lifespan. *Journal for Educational Research Online*, 5, 80.

Neumann, M., Becker, M., Maaz, K. (2014). Soziale Ungleichheiten in der Kompetenzentwicklung in der Grundschule und der Sekundarstufe I. *Zeitschrift für Erziehungswissenschaft*, 17, 167–203.

Passaretta, G., Skopek, J.,van Huizen, T. (2022). Is social inequality in school-age achievement generated before or during schooling? A European perspective. *European Sociological Review*, jcac005.

Pfeffer, F. T. (2008). Persistent inequality in educational attainment and its institutional context. *European Sociological Review*, 24, 543–565.

Pfost, M., Dörfler, T., Artelt, C. (2012). Reading competence development of poor readers in a German elementary school sample: An empirical examination of the Matthew effect model. *Journal of Research in Reading*, 35, 411–426.

Pfost, M., Hattie, J., Dörfler, T., 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, 203–244.

Pohl, S., Carstensen, C. H. (2012). NEPS Technical Report – Scaling the Data of the Competence Tests. *NEPS Working Papers No. 14*.

Ready, D. D. (2013). Associations between student achievement and student learning: Implications for value-added school accountability models. *Educational Policy*, 27, 92–120.

Relikowski, I., Schneider, T., Lineberg, T. (2015). Rezeptive Wortschatz- und Grammatikkompetenzen von Fünfjährigen mit und ohne Migrationshintergrund. *Frühe Bildung*, 4, 135–143.

Ritchie, S. J., Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological science*, 24, 1301–1308.

Schnabel, K. U., Alfeld, C., Eccles, J. S., Köller, O., Baumert, J. (2002). Parental influence on students' educational choices in the United States and Germany: Different ramifications – Same effect? *Journal of Vocational Behavior*, 60, 178–198.

Schneider, T. (2013). School class composition and student development in cognitive and non-cognitive domains: Longitudinal analyses of primary school students in Germany. In Windzio, M. (Ed.), *Integration and Inequality in Educational Institutions*. Dordrecht: Springer, pp. 167–190.

Skopek, J., Passaretta, G. (2021). Socioeconomic inequality in children's achievement from infancy to adolescence: The case of Germany. *Social Forces*, 100, 86–112.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 360–407.

Von Hippel, P. T., Workman, J., Downey, D. B. (2018). Inequality in reading and math skills forms mainly before kindergarten: A replication, and partial correction, of "Are schools the great equalizer?" *Sociology of Education*, 91, 323–357.

Von Hippel, P. T., Hamrock, C. (2019). Do test score gaps grow before, during, or between the school years? Measurement artifacts and what we can know in spite of them. *Sociological Science*, 6, 43–80.

Data Availability Statement

The data underlying this article were accessed from the NEPS Data Center [https://www.neps-data.de/Data-Center]; SC2: dx.doi.org/10.5157/NEPS:SC2:9.0.0; SC3: dx.doi.org/10.5157/NEPS:SC3:11.0.1; SC4: dx.doi.org/10.5157/NEPS:SC4:12.0.0). All analyses were run in the RemoteNEPS environment. Before you can access this environment, you need to sign a data usage contract and the RemoteNEPS Supplemental Agreement. (https://www.nepsdata.de/Data-Center/Data-Access/RemoteNEPS). The replication code (Stata do-files) is available in the Additional Online Material (replication_package.zip).

Table 1: Previous research from Germany on prior achievement and/or SES on achievement growth

Study	Data	N	Grades	Domain	Methods	PA → AG	SES → AG	Notes
Lehmann et al. (2001)	LAU (Hamburg)	11,849	5–6	Mathematics	Bivariate statistics	Compensation effects	–	
Schnabel et al. (2002)	BIJU (NRW and Berlin)	1,755	7–10	Mathematics	RVR	Compensation effects*	Positive SES effects	
Becker et al. (2006)	TIMMS (West German students, excluding students attending *Gesamtschule*)	1,864	7–8	Mathematics	Latent change models	Matthew effects	–	Only two measurement points; cautious interpretation of the positive correlation between intercept and slope
Ditton and Krüsken (2009)	KOALA-S (Bavaria and Saxony)	1,247	2–4	Mathematics and reading	Bivariate statistics and RVR	Compensation effects in mathematics and reading*	Positive SES effects in both domains	
Bos and Scharenberg (2010)	KESS (Hamburg)	8,774 (Grades 5–6) 5,951 (Grades 7–8)	5–6 7–8	Mathematics and reading	Multilevel RVR	Compensation effects in mathematics and reading*	Positive SES effects in both domains	
Baumert et al. (2012)	ELEMENT (Berlin)	3,167	4–6	Mathematics and reading	LGC	Compensation effects in reading; constant development in mathematics	Positive SES effects in both domains	Cumulative advantage in mathematics for students with high cognitive abilities
Pfost et al. (2012)	BiKS (Bavaria and Hesse)	1,124	3–4	Reading	LGC	Matthew effects	–	
Murayama et al. (2013)	PALMA (Bavaria)	3,530	5–10	Mathematics	LGC	Matthew effects	Stable SES differences	Zero correlation between intercept and slope once controlled for school track
Schneider (2013)	BiKS (Bavaria and Hesse)	2,379	3–4	Vocabulary and mathematics	Multilevel RVR	Compensation effects*	Positive SES effects in both domains	
Fleckenstein et al. (2019)	A longitudinal immersion study (Hamburg and Schleswig-Holstein)	590	1–4	Mathematics	LGC	Compensation effects	Stable SES differences	51.7% of the sample are immersion students
Neuendorf et al. (2020)	BiKS (Bavaria and Hesse, lower secondary students attending the Gymnasium)	1,010	5–9	Mathematics and reading	LGC	Compensation effects in both domains	–	Strong compensation between Grades 5 and 6; relatively constant thereafter

Skopek and Passaretta (2021)	NEPS (SC1, SC2, SC3) combining the different cohorts using an accelerated longitudinal design; national representative sample	16,512	Infancy–9	Mathematics, vocabulary, reading	Linear regression models	–	Relative SES gaps mainly open before schooling; slight increase over the course of primary school and secondary school	Absolute achievement measures: increase in SES gaps in math (primary school) and a slight increase in reading (lower secondary school)
Freund et al. (2021)	NEPS (SC4) (students tested in mathematics and reading in Grade 9); national representative sample	15,012	9–12 + 3 years	Mathematics and reading	LGC	Compensation effects	–	Negative intercept–slope correlation; up to 61% missing values in test scores (m = 30 imputations)
Passaretta et al. (2022)	NEPS (SC2)	420	Kindergarten–3	Vocabulary	RVR (using IV to account for measurement error)	–	Positive SES effects (stable SES differences using IV approach)	Relative achievement measures; zero effects using IV approach might be due to the low number of cases

Note: LGC = latent growth curves, RVR = regressor variable regressions (later achievement regressed on prior achievement), CS = change scores in achievement, AG = achievement growth, PA = prior achievement, NRW = North Rhine–Westphalia. * In the regression model (status attainment model), the coefficient of prior achievement on subsequent achievement was below 1.

Table 2: Correlation between prior achievement and achievement growth (between the first and last measurement time points)

	SC2	SC3	SC4
Domain			
Mathematics	-0.44	-0.37	-0.48
Language skills	-0.13	-0.54	-0.66

Source: NEPS SC2(9.0.0), SC3(11.0.1), SC4(12.0.0); based on 50 multiple-imputed data sets, author's calculations.

Table 3: Effects sizes (Cohen's delta) in achievement gaps between high and low SES students at selected time points

	SC2	SC2	SC3	SC3	SC4	SC4
	Grade 1	Grade 3/4	Grade 5	Grade 9	Grade 9	Grade 12
Domain						
Mathematics	0.89	1.00	0.95	1.00	0.29	0.30
Language skills	0.98	1.03	0.88	0.82	0.23	0.34

Source: NEPS SC2(9.0.0), SC3(11.0.1), SC4(12.0.0); based on 50 multiple-imputed data sets, author's calculations.

Figures

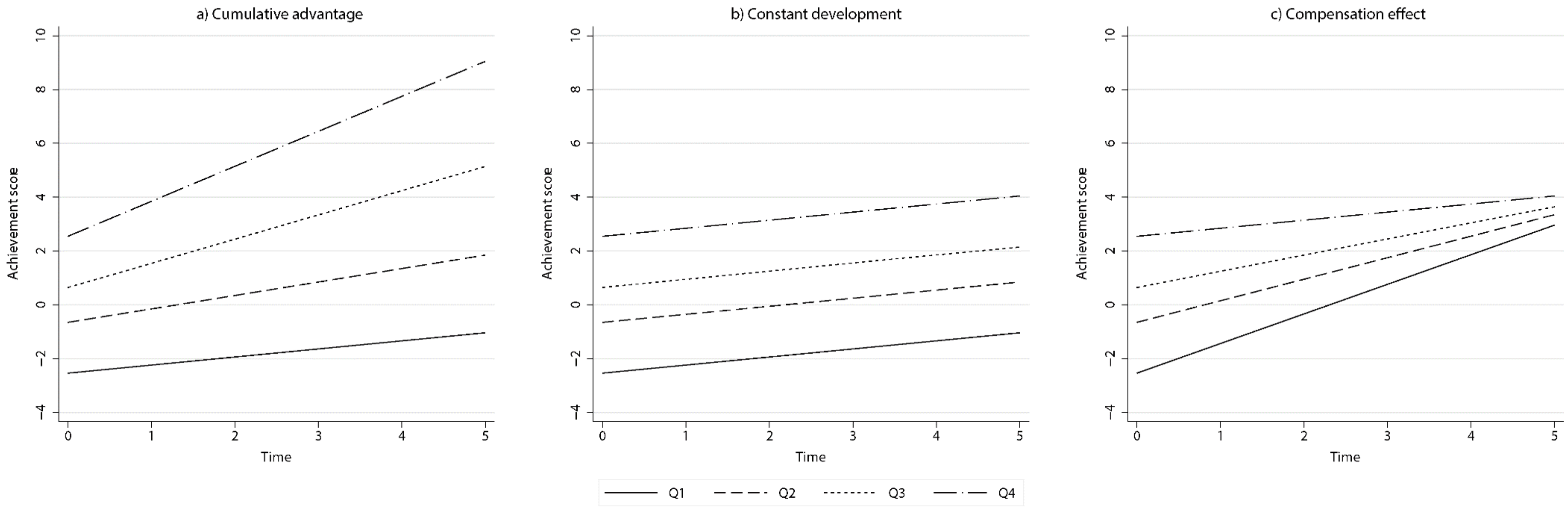

Figure 1: Possible achievement growth curves by achievement quartiles at first measurement point

Source: Author’s depiction.

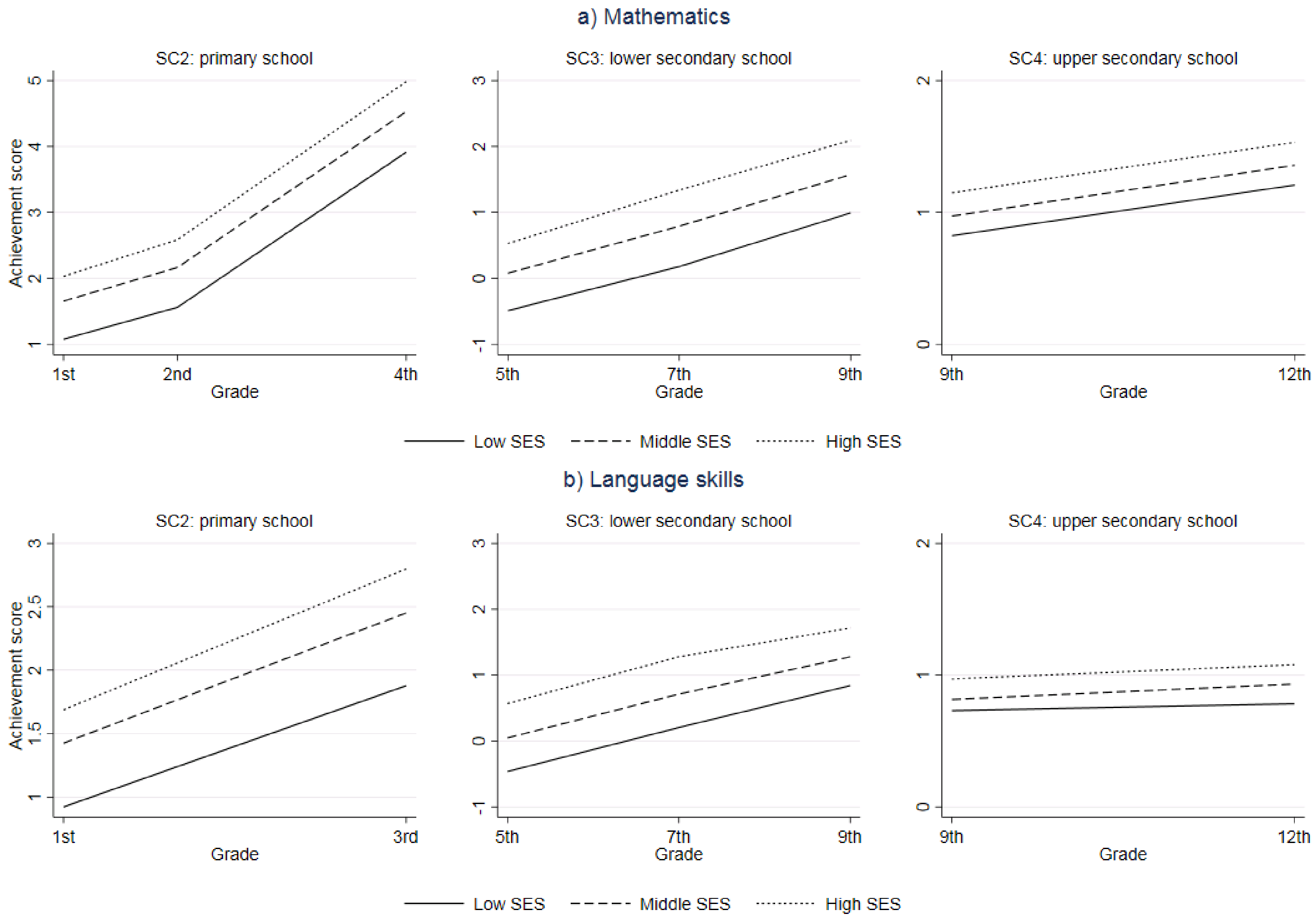

Figure 2: Achievement development over time by social origin, student cohort, and domain: a) mathematics, b) language skills

Source: NEPS SC2(9.0.0), SC3(11.0.1), SC4(12.0.0); based on 50 multiple-imputed data sets, author's calculations.

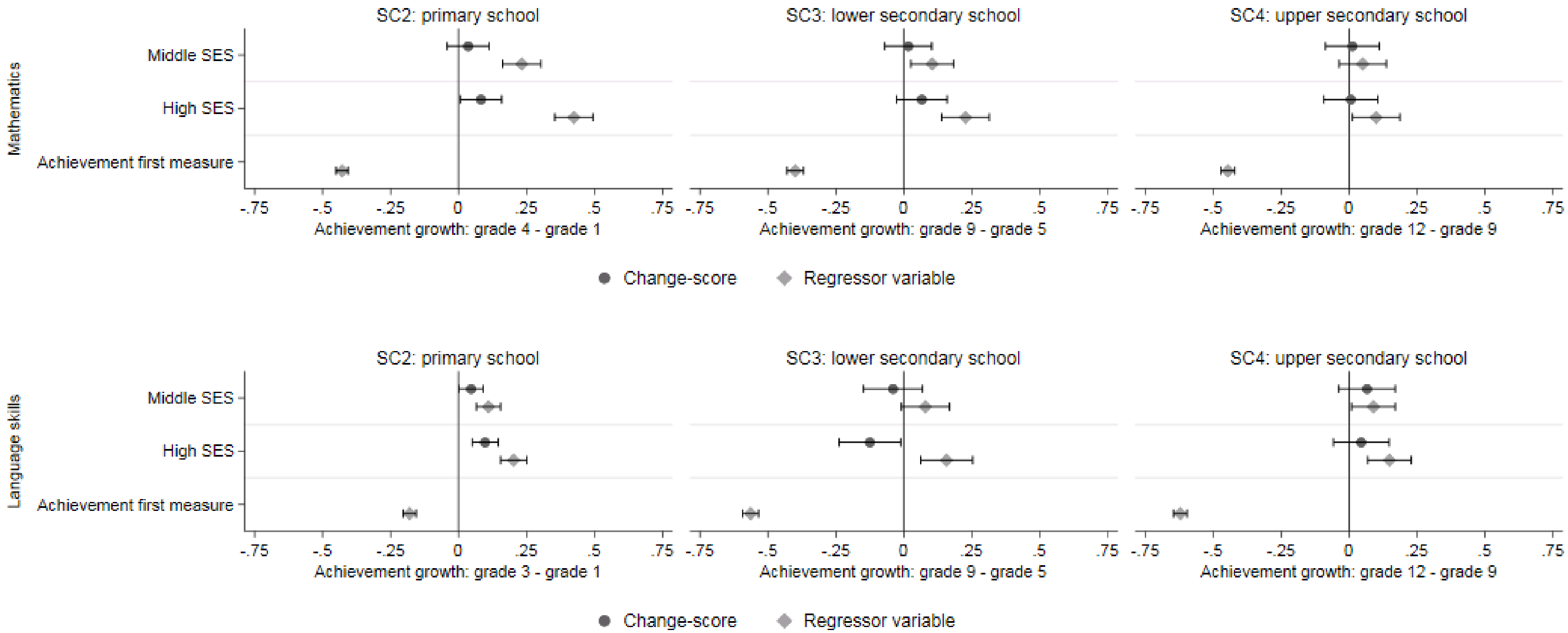

Figure 3: Prior achievement and SES coefficients on achievement growth (with 95% CI), by domain and student cohort

Note: Change score-model (based on Equation 5; unconditional on prior achievement). Regressor variable model (based on Equation 4; conditional on prior achievement). Full estimation results displayed in Tables C5 and C6 in the SOM.

Source: NEPS SC2(9.0.0), SC3(11.0.1), SC4(12.0.0); based on 50 multiple-imputed data sets, author's calculations.

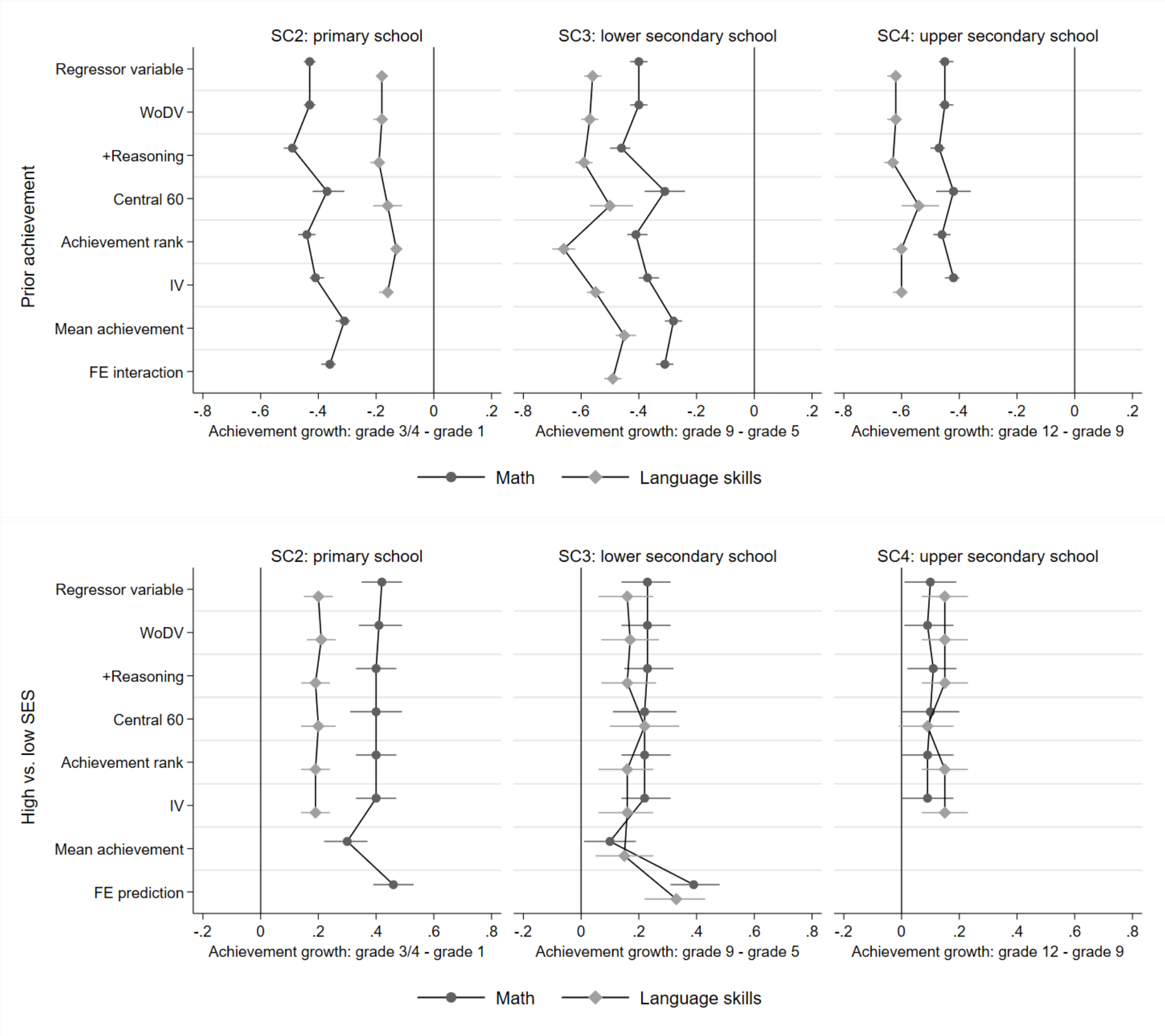

Figure 4: Robustness checks: prior achievement and SES coefficients (high SES vs low SES) on achievement growth (with 95% CI), by domain, student cohort, and estimation model

Note: Regressor-variable approach models only. Full estimation results displayed in Chapter D in the SOM. Mean achievement models and fixed effect models estimated only for cohorts with three measurement time points. Source: NEPS SC2(9.0.0), SC3(11.0.1), SC4(12.0.0); based on 50 multiple-imputed data sets, author’s calculations.

Supplemental material

No Matthew effects and stable SES gaps in achievement growth throughout schooling: Evidence from Germany. *European Sociological Review*

Contents

A The German School System

In comparison to other school systems, the German system is one of the most stratified (Bol et al., 2014). Usually, children start school at the age of six or seven, then attend primary school for four years. After primary school, the transition to the (mostly) tripartite secondary school system takes place. The three secondary tracks differ in terms of their requirements, length of school attendance, and attainable qualifications (Esser and Relikowski, 2015). First, the *Hauptschule* (the basic requirements secondary school) is the track with the lowest requirements, where a school leaving certificate is obtained at the end of Grade 9. Second, the *Realschule* (the extended requirements secondary school) has higher requirements and prepares its graduates for more demanding vocational training. At these schools, students receive their qualification at the end of Grade 10. Finally, the *Gymnasium* (grammar school) is the most demanding school track. With a degree (*Abitur*) acquired from this school after Grade 12, graduates are entitled to attend universities. Children from socially disadvantaged families are underrepresented in the most demanding track, even when controlling for academic ability (Neugebauer et al., 2013).

The 16 federal states are responsible for the organization of the school curricula. For this reason, some state school systems differ from those described above. For example, the duration of primary school may vary (in some states, primary school may last for six instead of four years), or there may be *Gesamtschulen* (comprehensive schools) or *Schulen mit mehreren Bildungsgängen* (secondary schools with several tracks) instead of *Hauptschulen* and *Realschulen*. In these lower secondary schools, there is no school tracking between schools but rather within schools. Access to the academic track also varies between the federal states in terms of how strongly school performance, teacher assessment, or parental will influence the transfer decision (Esser and Relikowski, 2015; Neugebauer et al., 2013).

B Achievement tests in the NEPS

B.1 Details on the achievement test

The achievement measurements (mathematics and language) in the NEPS are based on either an anchor-group or anchor-item design for successive measurement points (Pohl and Carstensen, 2012; Fischer et al., 2016). Since memory effects were expected for the competence tests in reading, an anchor-group design was used for the linking procedure of the competence scores onto a common scale. Since no memory effects were expected for the competence tests in mathematics, an anchor-item design was used for the linking procedure (Fischer et al., 2016).

To avoid memory effects due to the use of the same items over several waves, an anchor-group design was used for reading achievement. For this purpose, an 'independent link sample', which is not part of the actual study but comes from the same population as the respective starting cohort, was drawn. The competency tests of two consecutive waves were presented to this group at one point in time (Fischer et al., 2016: pp. 4, 10–11). Since the two tests have no items in common in the study group, they are each scaled independently. In the link sample, which processed both tests at the same time, all items of the two tests are scaled simultaneously (*concurrently*). This provides information about the item difficulty of the two tests and allows the determination of a correction parameter for the item difficulty of the items of the later test of the study group (for the exact procedure, see Fischer et al., 2016). The test scores of the study group in each wave were scaled independently for each grade level (Pohl and Carstensen, 2012). Then, the two test scores were linked by a linear transformation of the parameters of the second test, with the test scores of the first wave forming the reference scale (Fischer et al., 2016: p. 7).

Since no memory effects were expected for the mathematics achievement measurement, an anchor-item design was used. For this purpose, selected items that were used in the previously administered tests were reused in the subsequent tests. These common items were used to link the two tests on the basis of a common scale after checking certain preconditions (unidimensionality and measurement invariance). In addition, a correction term was considered to keep the item difficulties of the items included in both tests constant. Finally, the tests were linked using a linear transformation of the item parameter of the second test, by mean/mean-linking applying Rasch models (for the exact procedure, see Fischer et al., 2016).

Following the linking procedure, *weighted maximum likelihood estimates* (WLE estimators) were estimated (Pohl and Carstensen, 2012: p. 7). These WLE estimators represent point estimates for "the most likely competence score for each single person given the item responses of that person" (Pohl and Carstensen, 2012: p. 9). In this study, the uncorrected WLE estimators are used. These estimators allow examining skill development over time, "since differences in WLE scores can be interpreted as developmental trajectories across measurement points" (Fischer et al., 2016: p. 13). Hence, these WLE scores are suitable for longitudinal comparisons (e. g. Schnittjer et al., 2020; van de Ham et al., 2018; Fischer et al., 2017).

B.2 Test reliability

Table B1: Test reliability estimates of IRT scaled tests (WLE) by cohorts and wave

Competence Domain	Cohort	Wave	Grade	WLE reliability	Source
Mathematics	SC2	3	1	0.739	Schnittjer and Fischer, 2018
Mathematics	SC2	4	2	0.787	Schnittjer and Gerken, 2018
Mathematics	SC2	6	4	0.727	Schnittjer et al., 2020
Mathematics	SC3	1	5	0.778	Duchhardt and Gerdes, 2012
Mathematics	SC3	3	7	0.721	Schnittjer and Gerken, 2017
Mathematics	SC3	5	9	0.812	van de Ham et al., 2018
Mathematics	SC4	1	9	0.794	Duchhardt and Gerdes, 2013
Mathematics	SC4	7	12	0.766	Fischer et al., 2017
Receptive Vocabulary	SC2	3	1	0.87	Fischer and Durda, 2020
Receptive Vocabulary	SC2	5	3	0.84	Fischer and Durda, 2020
Reading	SC3	1	5	0.767	Pohl et al., 2012
Reading	SC3	3	7	0.791	Krannich et al., 2017
Reading	SC3	6	9	0.787	Scharl et al., 2017
Reading	SC4	2	9	0.749	Haberkorn et al., 2012
Reading	SC4	7	12	0.795	Gnambs et al., 2017

C Supplementary tables and figures

C.1 Descriptive results

Table C1: Means (M), Standard Deviations (SD), and number of non-missing Observations (N) for the three students cohorts (mathematics sample, before multiple imputation)

	SC2			SC3			SC4		
	Primary school			Lower secondary school			Upper secondary school		
	M	SD	N	M	SD	N	M	SD	N
Growth in Achievement	2.89	1.01	(5453)	1.52	0.90	(3036)	0.38	0.93	(3357)
Achievement first measure	1.71	1.13	(6481)	0.13	1.15	(3293)	1.04	1.14	(3836)
Social origin									
Low SES (0/1)	0.18	0.38	(6102)	0.23	0.42	(3221)	0.13	0.33	(3700)
Middle SES (0/1)	0.40	0.49	(6102)	0.38	0.49	(3221)	0.38	0.48	(3700)
High SES (0/1)	0.43	0.49	(6102)	0.39	0.49	(3221)	0.50	0.50	(3700)
Women (0/1)	0.51	0.50	(6866)	0.49	0.50	(3461)	0.55	0.50	(3946)
Minority student 0/1	0.28	0.45	(6614)	0.26	0.44	(3366)	0.21	0.41	(3919)
Age first measure (in months)	84.99	4.84	(6866)	130.81	5.98	(3461)	178.56	5.69	(3946)
Distance btw. first and last measurement (in months)	32.18	1.54	(6866)	48.12	0.65	(3439)	36.17	0.96	(3946)
School type									
Basic req. school (0/1)				0.11	0.31	(3461)			
Basic and extended school (0/1)				0.08	0.27	(3461)			
Extended req. school (0/1)				0.24	0.43	(3461)			
Comprehensive school (0/1)				0.07	0.25	(3461)			
Grammar school (0/1)				0.51	0.50	(3461)			
Number of students			6866			3461			3946
Number of schools			374			171			137
Number of federal states			16			13			16

Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0), author's calculations.

Table C2: Means (M), Standard Deviations (SD), and number of non-missing Observations (N) for the three students cohorts (language sample, before multiple imputation)

	SC2 Primary school			SC3 Lower secondary school			SC4 Upper secondary school		
	M	SD	N	M	SD	N	M	SD	N
Growth in achievement	1.05	0.61	(5272)	1.21	1.07	(2782)	0.09	0.98	(3290)
Achievement first measure	1.44	0.84	(6462)	0.13	1.25	(3188)	0.88	1.05	(3753)
Social origin									
Low SES (0/1)	0.18	0.38	(6100)	0.24	0.43	(3066)	0.13	0.33	(3664)
Middle SES (0/1)	0.40	0.49	(6100)	0.38	0.48	(3066)	0.38	0.48	(3664)
High SES (0/1)	0.43	0.49	(6100)	0.38	0.49	(3066)	0.49	0.50	(3664)
Women (0/1)	0.51	0.50	(6865)	0.49	0.50	(3348)	0.55	0.50	(3908)
Minority student 0/1	0.28	0.45	(6612)	0.26	0.44	(3258)	0.21	0.41	(3886)
Age first measure (in months)	84.98	4.84	(6865)	130.79	5.96	(3350)	178.55	5.69	(3908)
Distance btw. first and last measurement (in months)	20.25	1.82	(6490)	53.32	0.85	(3305)	30.53	0.88	(3908)
School type									
Basic req. school (0/1)				0.11	0.31	(3350)			
Basic and extended school (0/1)				0.08	0.27	(3350)			
Extended req. school				0.24	0.43	(3350)			
Comprehensive school (0/1)				0.07	0.25	(3350)			
Grammar school (0/1				0.51	0.50	(3350)			
Number of students			6865			3350			3908
Number of schools			374			172			137
Number of federal states			16			13			16

Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0), author's calculations.

Table C3: Means (M) and Standard Errors (SE) for the three students cohorts (mathematics sample, after multiple imputation)

	SC2 Primary school		SC3 Lower secondary school		SC4 Upper secondary school	
	M	SE	M	SE	M	SE
Growth in achievement	2.90	0.01	1.52	0.02	0.38	0.02
Achievement first measure	1.70	0.01	0.11	0.02	1.04	0.02
Social origin						
Low SES (0/1)	0.19	0.01	0.24	0.01	0.13	0.01
Middle SES (0/1)	0.40	0.01	0.38	0.01	0.38	0.01
High SES (0/1)	0.41	0.01	0.38	0.01	0.49	0.01
Women (0/1)	0.51	0.01	0.49	0.01	0.55	0.01
Minority student (0/1)	0.29	0.01	0.26	0.01	0.21	0.01
Age first measure (in months)	84.99	0.06	130.81	0.10	178.56	0.09
Distance btw. first and last measurement (in months)	32.18	0.02	48.12	0.01	36.17	0.02
School type						
Basic req. school (0/1)			0.11	0.01		
Basic and extended school (0/1)			0.08	0.00		
Extended req. school (0/1)			0.24	0.01		
Comprehensive school (0/1)			0.07	0.00		
Grammar school (0/1)			0.51	0.01		
Number of students		6866		3461		3946
Number of schools		374		171		137
Number of federal states		16		13		16

Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0). Estimates based on 50 multiple-imputed datasets, author's calculations.

Table C4: Means (M) and Standard Errors (SE) for the three students cohorts (language sample, after multiple imputation)

	SC2		SC3		SC4	
	Primary school		Lower secondary school		Upper secondary school	
	M	SE	M	SE	M	SE
Growth in achievement	1.05	0.01	1.22	0.02	0.10	0.02
Achievement first measure	1.44	0.01	0.12	0.02	0.88	0.02
Social origin						
Low SES (0/1)	0.19	0.01	0.24	0.01	0.13	0.01
Middle SES (0/1)	0.40	0.01	0.38	0.01	0.38	0.01
High SES (0/1)	0.41	0.01	0.38	0.01	0.49	0.01
Women (0/1)	0.51	0.01	0.49	0.01	0.55	0.01
Minority student (0/1)	0.29	0.01	0.27	0.01	0.21	0.01
Age first measure (in months)	84.98	0.06	130.79	0.10	178.55	0.09
Distance btw. first and last measurement (in months)	20.25	0.02	53.32	0.01	30.53	0.01
School type						
Basic req. school (0/1)			0.11	0.01		
Basic and extended school (0/1)			0.08	0.00		
Extended req. school (0/1)			0.24	0.01		
Comprehensive school (0/1)			0.07	0.00		
Grammar school (0/1)			0.51	0.01		
Number of students		6865		3350		3908
Number of schools		374		172		137
Number of federal states		16		13		16

Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0). Estimates based on 50 multiple-imputed datasets, author's calculations.

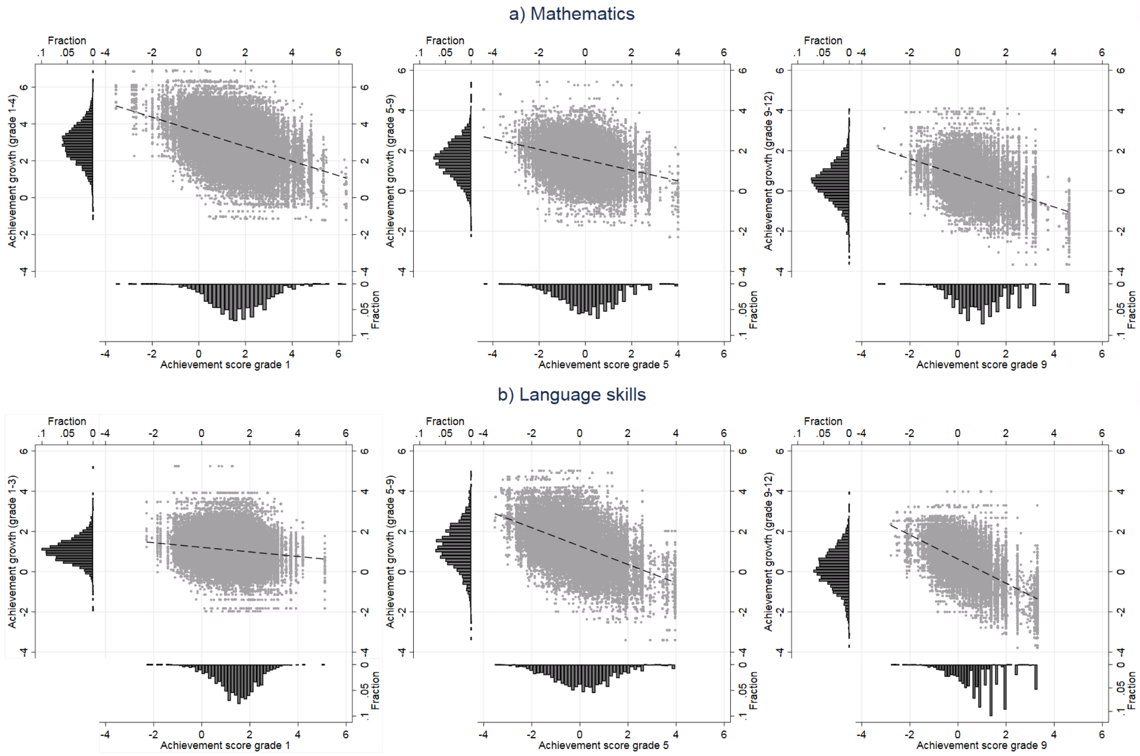

Figure C1: Association between prior achievement and subsequent achievement growth
Note: Dotted line = linear prediction based on regression from achievement growth on achievement score. Source: Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0). Based on 50 pooled multiple-imputed datasets, author's calculations.

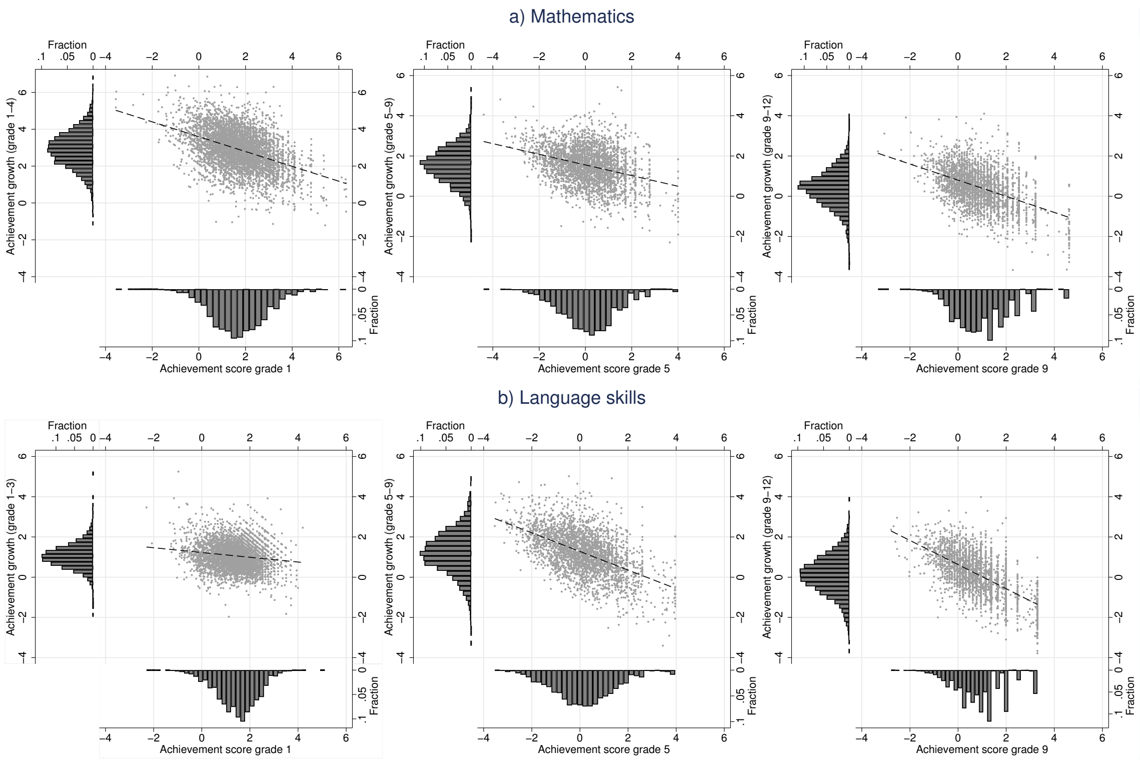

Figure C2: Association between prior achievement and subsequent achievement growth
Note: Dotted line = linear prediction based on regression from achievement growth on achievement score. Source: Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0). Based on complete cases only, author's calculations.

C.2 Multilevel results

Table C5: Multilevel models predicting achievement growth in mathematics for three student cohorts

	Grades 1–4		Grades 5–9		Grades 5–9		Grades 9–12	
	CS	RV	CS	RV	CS	RV	CS	RV
	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate
Social origin (ref: low SES)								
Middle SES (0/1)	0.04	0.23***	0.03	0.08	0.02	0.10**	0.01	0.05
	[-0.04,0.11]	[0.16,0.30]	[-0.06,0.12]	[-0.00,0.16]	[-0.07,0.10]	[0.03,0.18]	[-0.09,0.11]	[-0.04,0.14]
High SES (0/1)	0.08*	0.42***	0.08	0.17***	0.07	0.23***	0.01	0.10*
	[0.01,0.16]	[0.35,0.49]	[-0.01,0.18]	[0.09,0.26]	[-0.03,0.16]	[0.14,0.31]	[-0.09,0.11]	[0.01,0.19]
Achievement $t-1$ (c)		-0.43***		-0.45***		-0.40***		-0.45***
		[-0.45,-0.41]		[-0.48,-0.41]		[-0.43,-0.37]		[-0.47,-0.42]
Girl (0/1)	0.09***	-0.01	-0.02	-0.17***	-0.02	-0.15***	-0.02	-0.29***
	[0.04,0.14]	[-0.05,0.04]	[-0.08,0.04]	[-0.23,-0.11]	[-0.09,0.04]	[-0.21,-0.09]	[-0.08,0.04]	[-0.34,-0.23]
Minority student (0/1)	0.11***	-0.02	0.03	-0.09**	0.03	-0.08*	0.03	-0.09**
	[0.05,0.17]	[-0.08,0.03]	[-0.05,0.10]	[-0.16,-0.02]	[-0.05,0.10]	[-0.15,-0.01]	[-0.04,0.11]	[-0.16,-0.02]
Age first measure (in months) (c)	-0.03***	-0.02***	-0.00	-0.01**	-0.00	-0.01***	-0.01**	-0.01***
	[-0.03,-0.02]	[-0.03,-0.02]	[-0.01,0.00]	[-0.01,-0.00]	[-0.01,0.00]	[-0.02,-0.01]	[-0.01,-0.00]	[-0.02,-0.01]
Months btw. measurements (c)	0.10***	0.05***	-0.01	-0.02	-0.01	0.02	-0.01	-0.01
	[0.08,0.13]	[0.03,0.07]	[-0.08,0.06]	[-0.08,0.05]	[-0.08,0.06]	[-0.06,0.10]	[-0.05,0.03]	[-0.04,0.03]
School tracks (ref: Academic track)								
Basic req. (0/1)			0.15*	-0.62***				
			[0.01,0.29]	[-0.77,-0.47]				
Basic and extended req. (0/1)			0.06	-0.44***				
			[-0.13,0.24]	[-0.63,-0.26]				
Extended req. (0/1)			-0.07	-0.46***				
			[-0.18,0.04]	[-0.57,-0.35]				
Comprehensive (0/1)			0.02	-0.52***				
			[-0.16,0.21]	[-0.70,-0.34]				
Intercept	2.55***	2.47***	1.48***	1.71***	1.49***	1.44***	0.72***	0.63***
	[2.36,2.73]	[2.32,2.62]	[1.25,1.71]	[1.49,1.94]	[1.26,1.72]	[1.17,1.70]	[0.49,0.94]	[0.44,0.83]
SD(school)	0.24***	0.18***	0.17***	0.18***	0.18***	0.26***	0.11***	0.09***
	[0.21,0.28]	[0.15,0.22]	[0.13,0.22]	[0.14,0.22]	[0.14,0.23]	[0.22,0.32]	[0.07,0.17]	[0.06,0.15]
SD(student)	0.95***	0.85***	0.87***	0.78***	0.87***	0.78***	0.91***	0.78***
	[0.93,0.97]	[0.84,0.87]	[0.85,0.90]	[0.76,0.80]	[0.85,0.90]	[0.76,0.81]	[0.88,0.93]	[0.76,0.80]
No. of students	6866	6866	3461	3461	3461	3461	3946	3946
No. of schools	374	374	171	171	171	171	137	137
No. of federal states	16	16	13	13	13	13	16	16

Note: CS = change score modles. RV = regressor-variable models. 95% confidence intervals in brackets. (c) = Variables are mean centered. Estimates based on 50 multiple-imputed datasets. Federal state fixed effects included (not shown for data protection reasons). Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0), author's calculations.

Table C6: Multilevel models predicting achievement growth in language skills for three student cohorts

	Grades 1–3		Grades 5–9		Grades 5–9		Grades 9–12	
	CS	RV	CS	RV	CS	RV	CS	RV
	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate
Social origin (ref: low SES)								
Middle SES (0/1)	0.05*	0.11***	-0.02	0.05	-0.04	0.08	0.07	0.09*
	[0.00,0.09]	[0.07,0.15]	[-0.13,0.09]	[-0.04,0.13]	[-0.15,0.07]	[-0.01,0.17]	[-0.04,0.17]	[0.01,0.17]
High SES (0/1)	0.10***	0.20***	-0.09	0.09	-0.12*	0.16**	0.04	0.15***
	[0.05,0.15]	[0.15,0.25]	[-0.21,0.02]	[-0.00,0.19]	[-0.24,-0.01]	[0.06,0.25]	[-0.06,0.15]	[0.07,0.23]
Achievement $t-1$ (c)		-0.18***		-0.60***		-0.56***		-0.62***
		[-0.20,-0.16]		[-0.63,-0.57]		[-0.59,-0.53]		[-0.65,-0.60]
Girl (0/1)	-0.03	-0.04*	0.05	0.10**	0.05	0.11***	-0.06	0.09**
	[-0.06,0.01]	[-0.07,-0.00]	[-0.03,0.13]	[0.04,0.17]	[-0.03,0.13]	[0.05,0.17]	[-0.13,0.00]	[0.03,0.14]
Minority student (0/1)	-0.05**	-0.15***	0.08	-0.08	0.09	-0.06	0.04	-0.06
	[-0.09,-0.02]	[-0.19,-0.11]	[-0.02,0.17]	[-0.16,0.00]	[-0.01,0.18]	[-0.14,0.02]	[-0.04,0.12]	[-0.12,0.01]
Age first measure (in months) (c)	-0.01***	-0.01**	0.00	-0.00	0.00	-0.01	-0.00	-0.01***
	[-0.01,-0.00]	[-0.01,-0.00]	[-0.01,0.01]	[-0.01,0.00]	[-0.00,0.01]	[-0.01,0.00]	[-0.01,0.00]	[-0.02,-0.01]
Months btw. measurements (c)	0.02**	0.02**	-0.01	0.01	-0.01	0.01	0.02	0.02
	[0.01,0.03]	[0.00,0.03]	[-0.07,0.05]	[-0.05,0.06]	[-0.07,0.05]	[-0.06,0.09]	[-0.03,0.06]	[-0.02,0.06]
School tracks (ref: Academic track)								
Basic req. (0/1)			0.15	-0.80***				
			[-0.01,0.31]	[-0.95,-0.65]				
Basic and extended req. (0/1)			0.11	-0.49***				
			[-0.10,0.31]	[-0.68,-0.30]				
Extended req. (0/1)			0.03	-0.42***				
			[-0.09,0.15]	[-0.53,-0.31]				
Comprehensive (0/1)			0.23*	-0.42***				
			[0.03,0.43]	[-0.60,-0.24]				
Intercept	1.10***	1.08***	1.50***	1.58***	1.58***	1.33***	-0.02	-0.15
	[1.00,1.21]	[0.98,1.18]	[1.24,1.77]	[1.34,1.81]	[1.32,1.84]	[1.03,1.63]	[-0.24,0.21]	[-0.35,0.04]
SD(school)	0.12***	0.11***	0.15***	0.16***	0.15***	0.29***	0.10***	0.12***
	[0.10,0.15]	[0.09,0.14]	[0.10,0.22]	[0.12,0.22]	[0.10,0.23]	[0.24,0.35]	[0.05,0.18]	[0.09,0.16]
SD(student)	0.59***	0.57***	1.05***	0.84***	1.05***	0.84***	0.97**	0.73***
	[0.57,0.60]	[0.56,0.59]	[1.02,1.07]	[0.82,0.86]	[1.02,1.07]	[0.82,0.87]	[0.94,0.99]	[0.71,0.75]
No. of students	6865	6865	3350	3350	3350	3350	3908	3908
No. of schools	374	374	172	172	172	172	137	137
No. of federal states	16	16	13	13	13	13	16	16

Note: CS = change score modles. RV = regressor-variable models. 95% confidence intervals in brackets. (c) = Variables are mean centered. Estimates based on 50 multiple-imputed datasets. Federal state fixed effects included (not shown for data protection reasons). Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0), author's calculations.

C.2.1 Interpretation of the control variables

With respect to gender, the following effects emerge. In the models in which prior achievement is not taken into account (change score models), there are no gender differences in achievement growth for both domains in lower and upper secondary school. In elementary school, there are no significant differences in language skills for mathematics, but there is a significant advantage for girls. When controlling for prior achievement (regressor-variable models), the expected differences between the genders are found - especially in the secondary school cohorts - boys achieve significantly higher gains in mathematics and girls in language skills. In primary school, no gender differences are found in math achievement growth and small advantages for boys in vocabulary growth. These differences between the institutional settings could be due to gender roles becoming more pronounced with increasing age or due to the fact that especially in the higher grades of the lower secondary school and in the upper secondary school, courses in German and mathematics with different levels of requirements can be chosen.

Minority students show significantly lower language skill gains in primary school (in both change score and regressor-variable models). In lower and upper secondary school, minority students do not differ significanly in growth rates from majority students. For mathematics skill growth, once controlled for prior achievement, nor differences between minority and majority in mathematics skill development are evident in primary school. These differences become apparent in lower and upper secondary school: majority students have higher mathematics skill growth than minority students - given the same initial achievement.

For math skills, it is evident that the older the students were when first measured, the smaller the gains in all three institutional settings. Negative effects are also found in language skills in all three institutional settings, but these are only significant in primary school and upper secondary school. The time between measurement points (months between individual measurements) only has statistically significant positive effects on growth rates in primary school in both mathematic skills and language skills.

Looking at the random effects (SD(school)) across all models shows that these are statistically significantly different from 0 and that the school level can thus contribute to the elucidation of growth differences.

C.3 Additional multilevel results

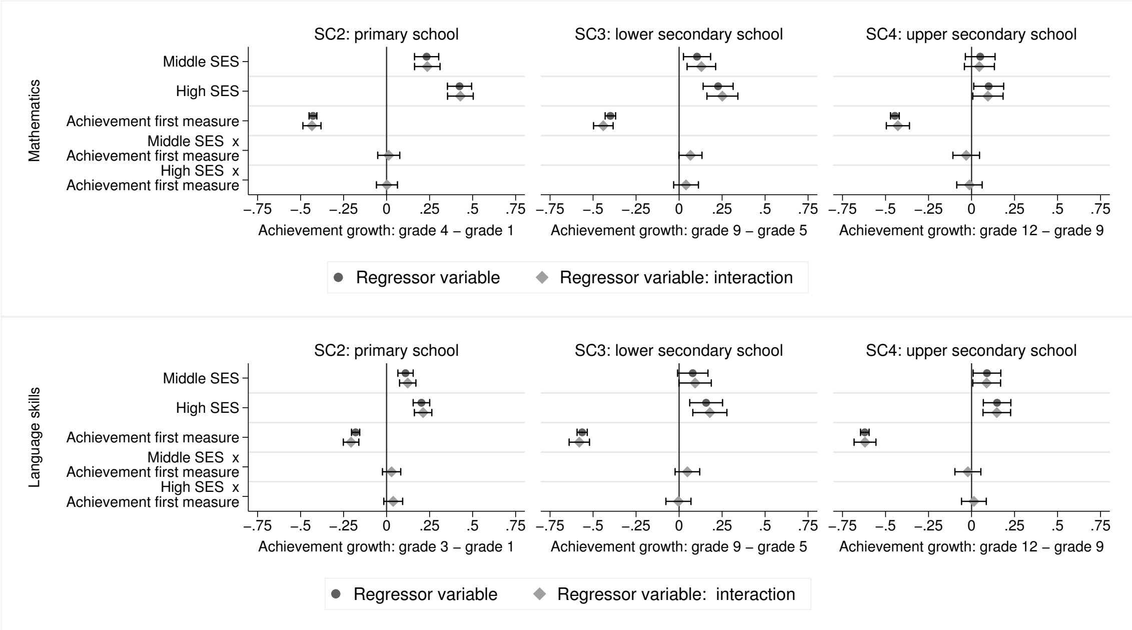

Figure C3: Prior achievement and SES and their interaction coefficients on achievement growth (with 95% CI), by domain and student cohort
Note: Change score model (based on Equation 5; unconditional on prior achievement). Regressor variable model (based on Equation 4; conditional on prior achievement). Source: Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0). Based on 50 multiple-imputed datasets, author's calculations.

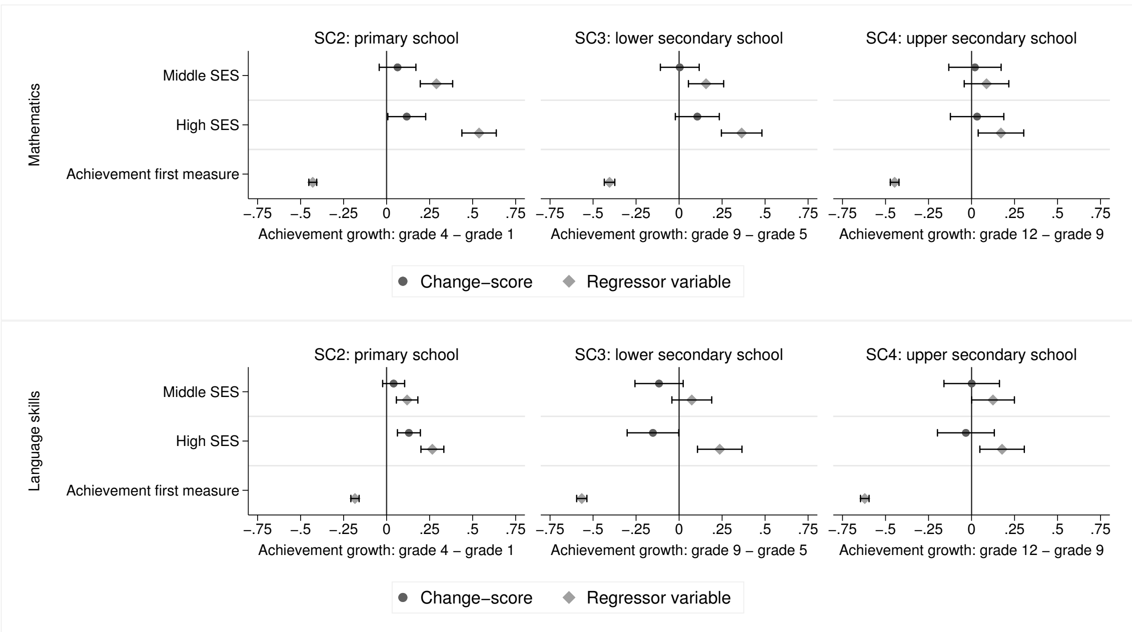

Figure C4: Prior achievement and SES (measured using parental CASMIN) coefficients on achievement growth (with 95% CI), by domain and student cohort
Note: Change score model (based on Equation 5; unconditional on prior achievement). Regressor variable model (based on Equation 4; conditional on prior achievement). Low SES = CASMIN (1a-c), middle SES = CASMIN (2a-c) and high SES = CASMIN (3a+3b).
Source: Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0). Based on 50 multiple-imputed datasets, author's calculations.

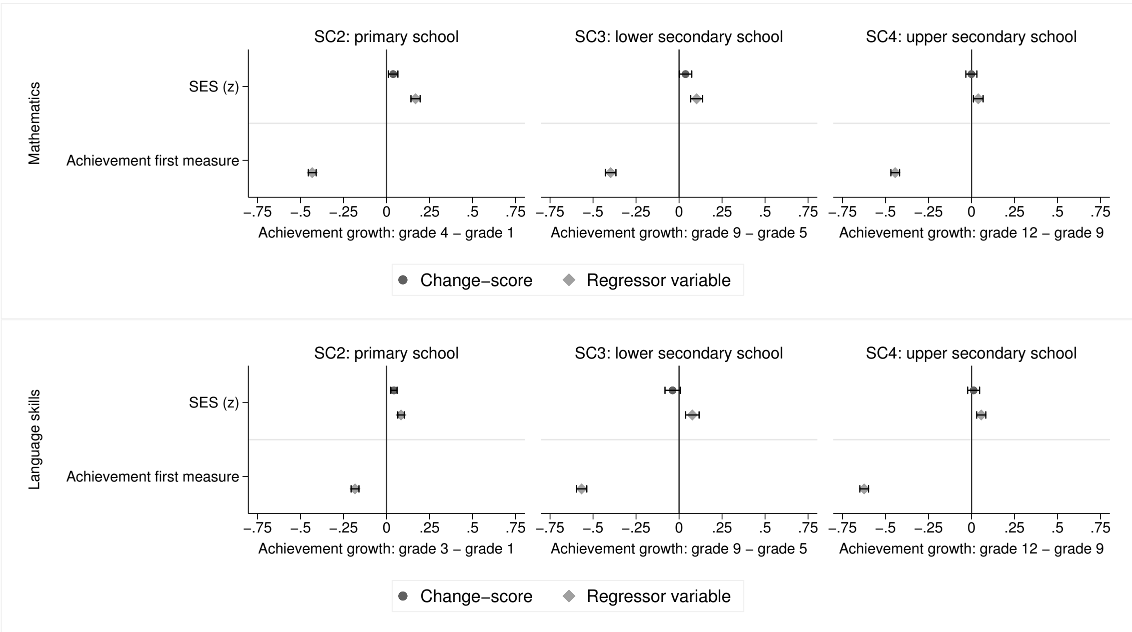

Figure C5: Prior achievement and SES (measured using z-standardized ISEI) coefficients on achievement growth (with 95% CI), by domain and student cohort

Note: Change score model (based on Equation 5; unconditional on prior achievement). Regressor variable model (based on Equation 4; conditional on prior achievement). Source: Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0). Based on 50 multiple-imputed datasets, author's calculations.

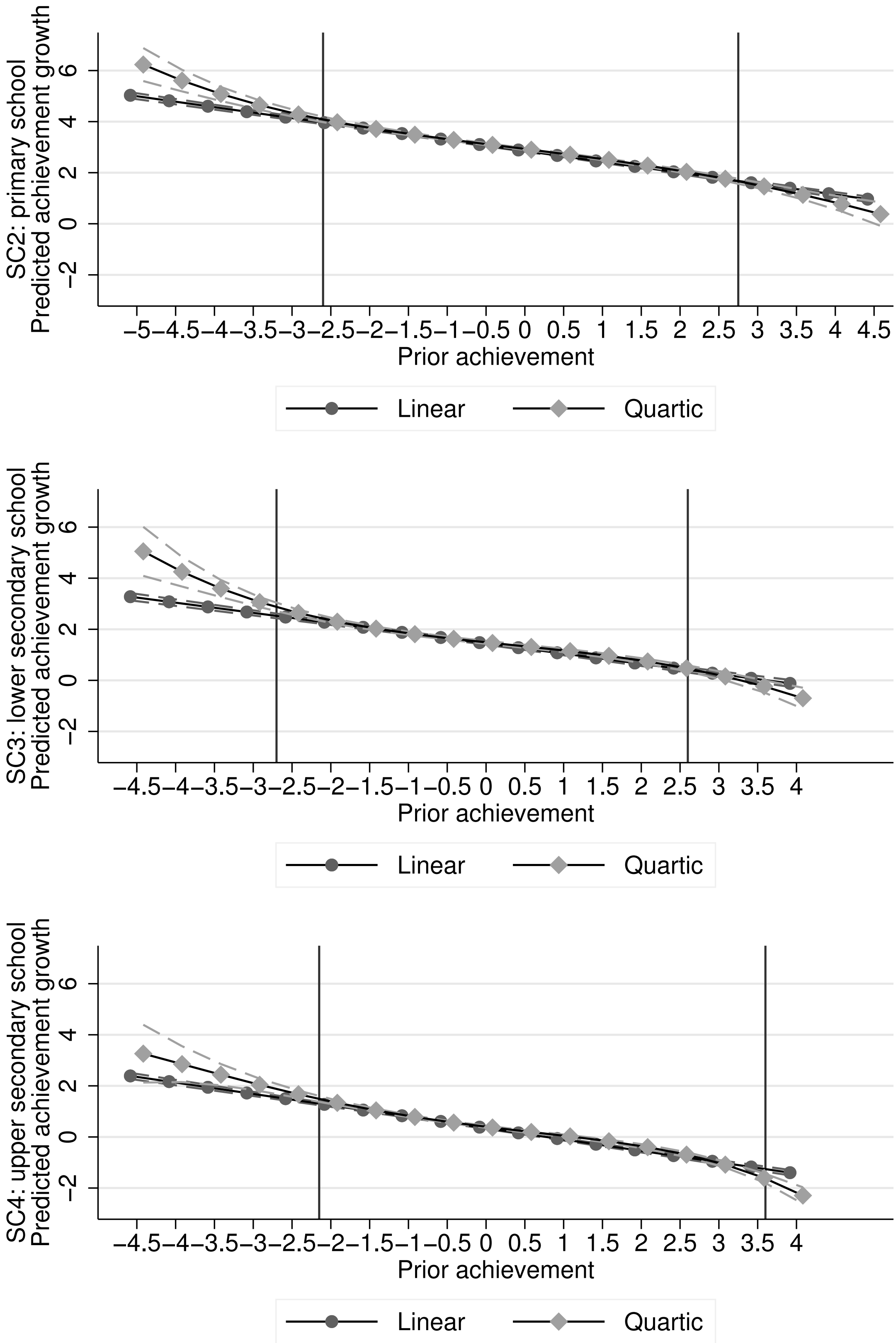

Figure C6: Predective achievement growth (with 95% CI) in math over initial achievement using linear and quartic growth specifications, by domain and student cohort

Predictive margins (fixed part only) are depicted. For the predictions all other model variables were kept at the mean. Note: Values rang between higest and lowest observed data points in the data. The area within the vertical lines indicate the central 98% of the observed data points. Source: Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0). Based on 50 multiple-imputed datasets, author's calculations.

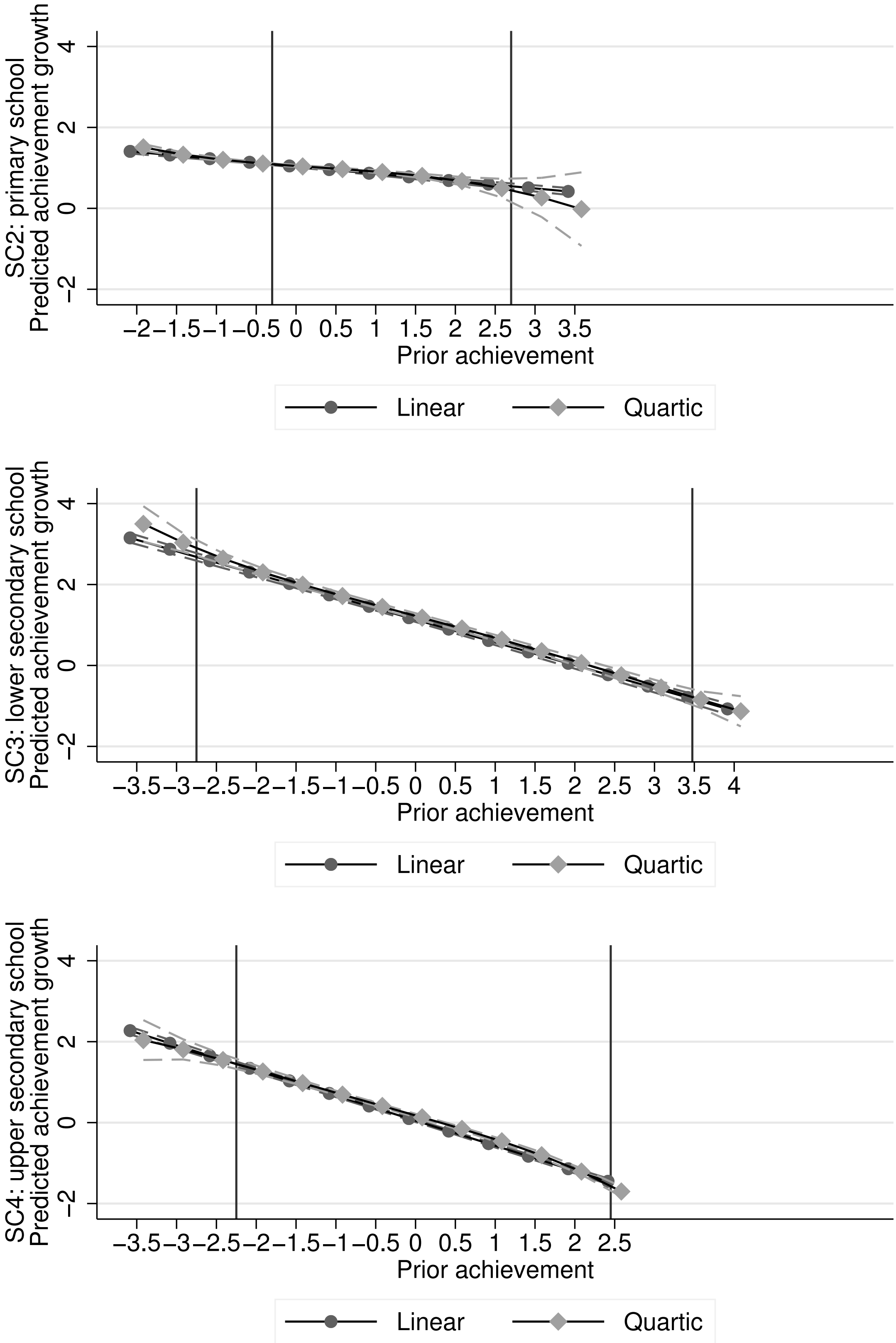

Figure C7: Predective achievement growth (with 95% CI) in language skills over initial achievement using linear and quartic growth specifications, by domain and student cohort
Predictive margins (fixed part only) are depicted. For the predictions all other model variables were kept at the mean. Note: Values rang between higest and lowest observed data points in the data. The area within the vertical lines indicate the central 98% of the observed data points. Source: Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0). Based on 50 multiple-imputed datasets, author's calculations.

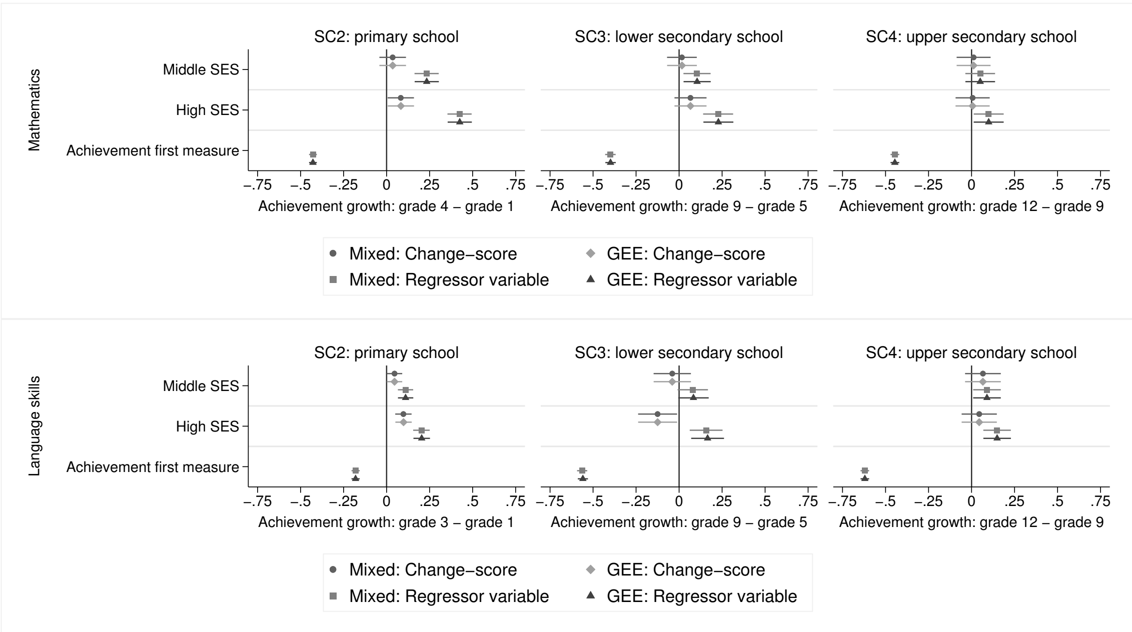

Figure C8: Prior achievement and SES coefficients on achievement growth (with 95% CI) using different statistical models to account for the clustering of the data, by domain and student cohort

Note: Change score model (based on Equation 5; unconditional on prior achievement). Regressor variable model (based on Equation 4; conditional on prior achievement). GEE models calculated using a gaussian link function and exchangable within-cluster correlations (calculated using the xtgee Stata command; schools as clusters).
Source: Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0). Based on 50 multiple-imputed datasets, author's calculations.

D Robustness checks

I have conducted a number of robustness checks. In the following subsections, I briefly explain why I conducted them, briefly describe my procedure, and present the results. As in the paper, I present the models of the regressor-variable approach and, where appropriate, the change-score approach. Finally, I have additionally estimated achievement growth using growth curve models for the cohorts with three achievement scores available (see subsection D.5). Since this represents a different statistical modeling approach than the one I use my paper, I present these results separately.

D.1 Ceiling effects and measurement sensitivity

As one can see in Figure D2 — especially in cohorts SC3 and SC4 — there are some very good students at the first achievement assessment who deteriorate considerably over time. This could be an indication of ceiling effects or a lack of sensitivity of the achievement assessments in the extreme ranges. Thus, the question arises whether the reported compensatory effects might be influenced by the incorrect recording of the growth of the high achievers. To examine this, I grouped the students into deciles based on their test scores and plotted the proportion of students who made achievement gains (see Figures D1 and D2).

Following the argument of Kelly and Yu (2017, p. 356), a ceiling effect is present when "students who "max out" the test at Time 1 have nowhere to go but down". Accordingly, a large proportion of high performing students should not be able to improve. This is not evident for the SC2 and SC3 starting cohorts. There, even among the best performing students, an majority can improve over time. In SC4, for mathematics the majority of students also improve except for the best decile. For reading, the majority of students also improve except for the top 40 percent. These results suggest rather that the sensitivity of the tests is lower at the peripheries than that there are strong ceiling effects (especially for SC2 and SC3). However, it is striking that the proportion of students with achievement growth is considerably lower in all deciles of SC4 than in the other two cohorts. I think two reasons may account for this pattern. First, many studies have been able to show that achievement growth decreases with increasing age. The period examined in SC4 covers the ages between 15 and 19. Second, this is a selective population with already very good performance levels in the most demanding school track (Dollmann, 2016; Schneider, 2008). To check whether the sensitivity of achievement tests affects my results, I estimated models in which I excluded the part of the sample belonging to the best or worst 20 percent of the first achievement measurement (see Tables D1 and D2).

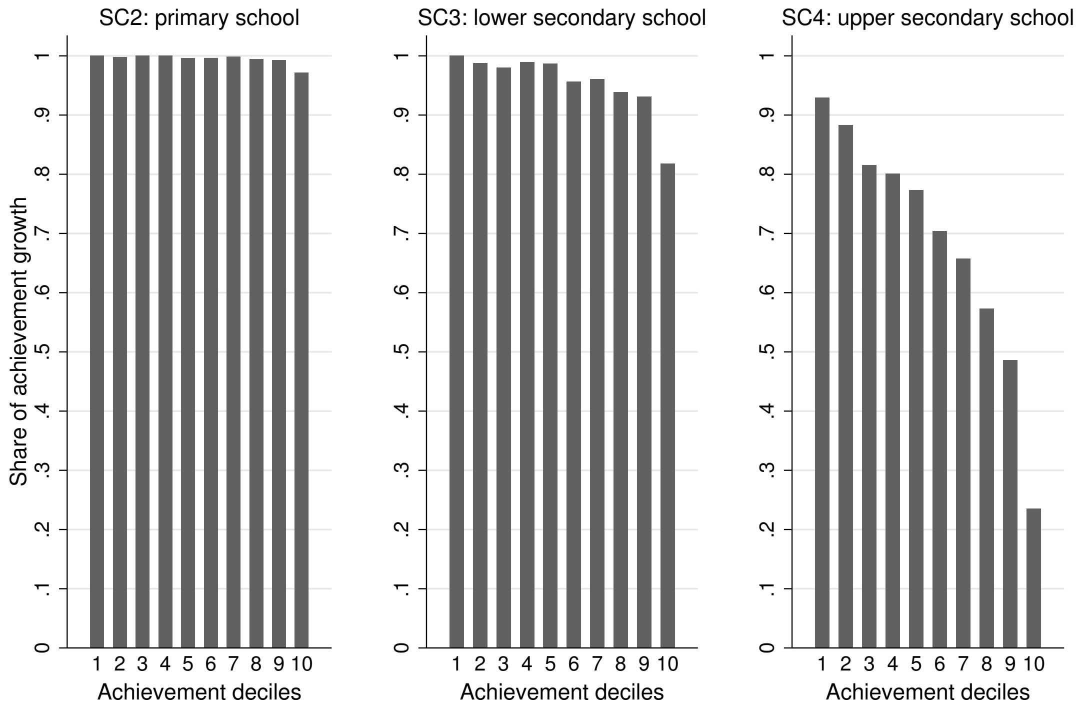

Figure D1: Share of positive achievement growth over time, by achievement decile and student cohort for mathematics
Source: Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0). Based on 50 multiple-imputed datasets, author's calculations.

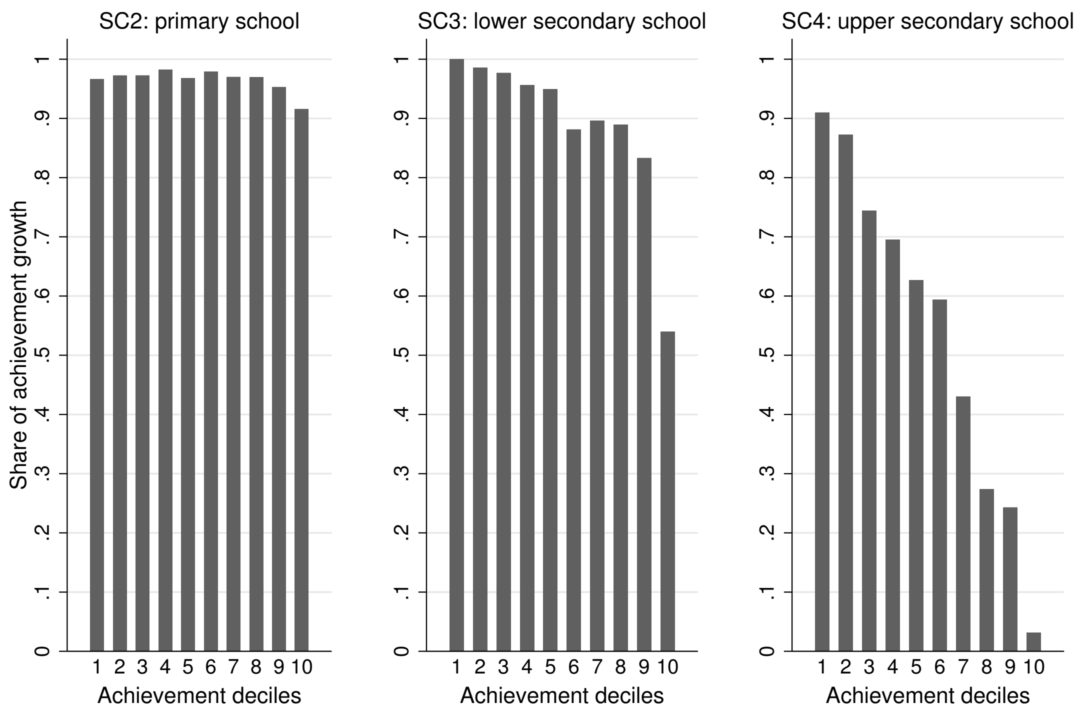

Figure D2: Share of positive achievement growth over time, by achievement decile and student cohort for language skills
Source: Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0). Based on 50 multiple-imputed datasets, author's calculations.

Table D1: Multilevel models predicting achievement growth in mathematics for three student cohorts (central 60 percent)

	Grades 1–4		Grades 5–9		Grades 5–9		Grades 9–12	
	CS	RV	CS	RV	CS	RV	CS	RV
	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate
Social origin (ref: low SES)								
Middle SES (0/1)	0.15***	0.21***	0.04	0.06	0.06	0.08	0.04	0.04
	[0.07,0.24]	[0.12,0.29]	[-0.06,0.14]	[-0.04,0.16]	[-0.04,0.17]	[-0.02,0.18]	[-0.07,0.15]	[-0.06,0.15]
High SES (0/1)	0.31***	0.40***	0.13*	0.16**	0.19***	0.22***	0.10	0.10
	[0.22,0.40]	[0.31,0.49]	[0.02,0.24]	[0.05,0.27]	[0.08,0.30]	[0.11,0.33]	[-0.01,0.20]	[-0.00,0.20]
Achievement $t-1$ (c)		-0.37***		-0.38***		-0.31***		-0.42***
		[-0.42,-0.31]		[-0.45,-0.30]		[-0.38,-0.24]		[-0.48,-0.36]
Girl (0/1)	0.01	-0.02	-0.15***	-0.19***	-0.14***	-0.17***	-0.21***	-0.28***
	[-0.05,0.07]	[-0.07,0.04]	[-0.22,-0.07]	[-0.26,-0.12]	[-0.21,-0.06]	[-0.24,-0.10]	[-0.28,-0.14]	[-0.35,-0.21]
Minority student (0/1)	-0.00	-0.04	-0.06	-0.11*	-0.05	-0.08	-0.09*	-0.11**
	[-0.07,0.07]	[-0.10,0.03]	[-0.15,0.03]	[-0.19,-0.02]	[-0.14,0.04]	[-0.17,0.01]	[-0.17,-0.00]	[-0.19,-0.03]
Age first measure (in months) (c)	-0.03***	-0.02***	-0.01*	-0.01*	-0.01**	-0.01***	-0.01***	-0.02***
	[-0.03,-0.02]	[-0.03,-0.02]	[-0.02,-0.00]	[-0.02,-0.00]	[-0.02,-0.00]	[-0.02,-0.00]	[-0.02,-0.01]	[-0.02,-0.01]
Months btw. measurements (c)	0.07***	0.06***	-0.02	-0.03	0.00	-0.00	-0.00	-0.01
	[0.05,0.10]	[0.04,0.09]	[-0.10,0.06]	[-0.11,0.05]	[-0.08,0.08]	[-0.09,0.09]	[-0.04,0.04]	[-0.05,0.03]
School tracks (ref: Academic track)								
Basic req. (0/1)			-0.36***	-0.55***				
			[-0.55,-0.17]	[-0.75,-0.36]				
Basic and extended req. (0/1)			-0.31**	-0.47***				
			[-0.53,-0.09]	[-0.69,-0.24]				
Extended req. (0/1)			-0.28***	-0.40***				
			[-0.41,-0.16]	[-0.53,-0.28]				
Comprehensive (0/1)			-0.25*	-0.42***				
			[-0.47,-0.02]	[-0.64,-0.19]				
Intercept	2.52***	2.51***	1.68***	1.74***	1.51***	1.48***	0.65***	0.58***
	[2.34,2.69]	[2.34,2.68]	[1.42,1.95]	[1.47,2.00]	[1.24,1.78]	[1.19,1.78]	[0.42,0.88]	[0.35,0.80]
SD(school)	0.16***	0.15***	0.19***	0.20***	0.23***	0.27***	0.09***	0.10***
	[0.11,0.22]	[0.11,0.21]	[0.15,0.25]	[0.15,0.26]	[0.18,0.29]	[0.21,0.33]	[0.04,0.17]	[0.05,0.17]
SD(student)	0.84***	0.82***	0.78***	0.76***	0.78***	0.76***	0.77***	0.74***
	[0.82,0.86]	[0.80,0.84]	[0.75,0.81]	[0.73,0.78]	[0.75,0.81]	[0.73,0.79]	[0.75,0.79]	[0.71,0.76]
No. of students	3888	3888	1975	1975	1975	1975	2409	2409
No. of schools	372	372	166	166	166	166	137	137
No. of federal states	16	16	13	13	13	13	16	16

Note: CS = change score modles. RV = regressor-variable models. 95% confidence intervals in brackets. (c) = Variables are mean centered. Estimates based on 50 multiple-imputed datasets. Federal state fixed effects included (not shown for data protection reasons). Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0), author's calculations.

Table D2: Multilevel models predicting achievement growth in language skills for three student cohorts (central 60 percent)

	Grades 1–3		Grades 5–9		Grades 5–9		Grades 9–12	
	CS	RV	CS	RV	CS	RV	CS	RV
	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate
Social origin (ref: low SES)								
Middle SES (0/1)	0.09**	0.11***	-0.01	0.02	0.03	0.07	0.00	0.02
	[0.03,0.15]	[0.05,0.17]	[-0.12,0.11]	[-0.08,0.13]	[-0.09,0.14]	[-0.04,0.18]	[-0.10,0.11]	[-0.08,0.12]
High SES (0/1)	0.17***	0.20***	0.07	0.13*	0.14*	0.22***	0.06	0.09
	[0.11,0.23]	[0.14,0.26]	[-0.06,0.19]	[0.01,0.25]	[0.02,0.26]	[0.10,0.34]	[-0.05,0.16]	[-0.01,0.18]
Achievement $t-1$ (c)		-0.16***		-0.56***		-0.50***		-0.54***
		[-0.21,-0.11]		[-0.63,-0.48]		[-0.57,-0.42]		[-0.60,-0.47]
Girl (0/1)	-0.04*	-0.04*	0.12**	0.12**	0.12**	0.12**	0.07*	0.09**
	[-0.08,-0.00]	[-0.08,-0.00]	[0.04,0.21]	[0.04,0.20]	[0.04,0.21]	[0.04,0.20]	[0.01,0.14]	[0.03,0.15]
Minority student (0/1)	-0.10***	-0.12***	-0.05	-0.08	-0.03	-0.07	-0.06	-0.06
	[-0.15,-0.06]	[-0.17,-0.07]	[-0.15,0.06]	[-0.19,0.02]	[-0.14,0.07]	[-0.17,0.04]	[-0.14,0.02]	[-0.14,0.02]
Age first measure (in months) (c)	-0.01***	-0.01**	0.00	0.00	-0.00	-0.00	-0.01**	-0.01***
	[-0.01,-0.00]	[-0.01,-0.00]	[-0.00,0.01]	[-0.01,0.01]	[-0.01,0.01]	[-0.01,0.00]	[-0.02,-0.00]	[-0.02,-0.00]
Months btw. measurements (c)	0.02**	0.02*	0.00	-0.00	0.01	0.01	0.01	0.01
	[0.00,0.03]	[0.00,0.03]	[-0.06,0.07]	[-0.07,0.06]	[-0.06,0.08]	[-0.07,0.09]	[-0.04,0.06]	[-0.04,0.05]
School tracks (ref: Academic track)								
Basic req. (0/1)			-0.43***	-0.74***				
			[-0.62,-0.24]	[-0.93,-0.55]				
Basic and extended req. (0/1)			-0.25*	-0.45***				
			[-0.47,-0.02]	[-0.67,-0.23]				
Extended req. (0/1)			-0.23***	-0.37***				
			[-0.35,-0.10]	[-0.49,-0.25]				
Comprehensive (0/1)			-0.10	-0.30*				
			[-0.33,0.14]	[-0.53,-0.07]				
Intercept	1.06***	1.06***	1.54***	1.49***	1.42***	1.29***	0.12	-0.01
	[0.95,1.18]	[0.95,1.17]	[1.25,1.83]	[1.21,1.77]	[1.13,1.71]	[0.97,1.61]	[-0.11,0.35]	[-0.23,0.22]
SD(school)	0.10***	0.10***	0.16***	0.17***	0.20***	0.26***	0.12***	0.12***
	[0.07,0.14]	[0.07,0.14]	[0.10,0.24]	[0.12,0.24]	[0.14,0.27]	[0.21,0.34]	[0.07,0.18]	[0.08,0.18]
SD(student)	0.55***	0.55***	0.88***	0.83***	0.88***	0.83***	0.75***	0.71***
	[0.53,0.56]	[0.53,0.56]	[0.85,0.91]	[0.80,0.86]	[0.85,0.91]	[0.80,0.86]	[0.73,0.78]	[0.69,0.73]
No. of students	3900	3900	1965	1965	1965	1965	2408	2408
No. of schools	372	372	168	168	168	168	137	137
No. of federal states	16	16	13	13	13	13	16	16

Note: CS = change score modles. RV = regressor-variable models. 95% confidence intervals in brackets. (c) = Variables are mean centered. Estimates based on 50 multiple-imputed datasets. Federal state fixed effects included (not shown for data protection reasons). Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0), author's calculations.

D.2 Models without cases with imputed achievement growth

Since there is a considerable number of missing values in the dependent variable (see Tables C1 and C2), one could argue that the relationship between prior achievement and subsequent growth might be biased by the imputed cases. Therefore, I re-estimated the models but without the cases in which the dependent variable (achievement growth) was imputed. The results are presented in Table D3 and Table D4.

Table D3: Multilevel models predicting achievement growth in mathematics for three student cohorts (non-missing dependent variable)

	Grades 1–4		Grades 5–9		Grades 5–9		Grades 9–12	
	CS	RV	CS	RV	CS	RV	CS	RV
	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate
Social origin (ref: low SES)								
Middle SES (0/1)	0.05	0.24***	0.03	0.09*	0.02	0.11**	0.01	0.05
	[-0.03,0.13]	[0.16,0.31]	[-0.06,0.12]	[0.01,0.17]	[-0.07,0.11]	[0.03,0.20]	[-0.09,0.11]	[-0.04,0.14]
High SES (0/1)	0.08*	0.41***	0.08	0.17***	0.07	0.23***	-0.02	0.09*
	[0.00,0.16]	[0.34,0.49]	[-0.02,0.18]	[0.09,0.26]	[-0.03,0.16]	[0.14,0.31]	[-0.12,0.09]	[0.01,0.18]
Achievement $t-1$ (c)		-0.43***		-0.45***		-0.40***		-0.45***
		[-0.45,-0.41]		[-0.48,-0.42]		[-0.43,-0.37]		[-0.47,-0.42]
Girl (0/1)	0.07**	-0.01	-0.03	-0.18***	-0.03	-0.16***	-0.02	-0.29***
	[0.02,0.13]	[-0.06,0.03]	[-0.09,0.03]	[-0.24,-0.12]	[-0.10,0.03]	[-0.22,-0.10]	[-0.09,0.04]	[-0.35,-0.24]
Minority student (0/1)	0.12***	-0.02	0.03	-0.09*	0.03	-0.08*	0.02	-0.09**
	[0.06,0.18]	[-0.07,0.04]	[-0.04,0.11]	[-0.16,-0.02]	[-0.05,0.11]	[-0.15,-0.00]	[-0.06,0.10]	[-0.16,-0.03]
Age first measure (in months) (c)	-0.03***	-0.02***	-0.00	-0.01***	-0.00	-0.01***	-0.01**	-0.01***
	[-0.03,-0.02]	[-0.03,-0.02]	[-0.01,0.00]	[-0.01,-0.00]	[-0.01,0.00]	[-0.02,-0.01]	[-0.01,-0.00]	[-0.02,-0.01]
Months btw. measurements (c)	0.10***	0.05***	-0.02	-0.03	-0.02	0.01	-0.02	-0.01
	[0.08,0.13]	[0.03,0.07]	[-0.09,0.05]	[-0.09,0.04]	[-0.09,0.05]	[-0.07,0.09]	[-0.06,0.03]	[-0.04,0.03]
School tracks (ref: Academic track)								
Basic req. (0/1)			0.16*	-0.62***				
			[0.01,0.30]	[-0.77,-0.47]				
Basic and extended req. (0/1)			0.07	-0.43***				
			[-0.13,0.26]	[-0.62,-0.24]				
Extended req. (0/1)			-0.08	-0.46***				
			[-0.19,0.04]	[-0.58,-0.35]				
Comprehensive (0/1)			0.01	-0.53***				
			[-0.18,0.20]	[-0.71,-0.34]				
Intercept	2.55***	2.47***	1.47***	1.70***	1.48***	1.43***	0.74***	0.64***
	[2.35,2.74]	[2.30,2.64]	[1.23,1.72]	[1.47,1.93]	[1.25,1.72]	[1.15,1.70]	[0.51,0.97]	[0.44,0.83]
SD(school)	0.26***	0.21***	0.19***	0.19***	0.20***	0.27***	0.12***	0.10***
	[0.22,0.30]	[0.18,0.25]	[0.14,0.24]	[0.16,0.24]	[0.16,0.25]	[0.23,0.33]	[0.08,0.18]	[0.07,0.15]
SD(student)	0.94***	0.85***	0.87***	0.78***	0.87***	0.78***	0.91***	0.78***
	[0.93,0.96]	[0.83,0.86]	[0.85,0.90]	[0.76,0.80]	[0.85,0.90]	[0.76,0.80]	[0.89,0.93]	[0.76,0.80]
No. of students	5453	5453	3036	3036	3036	3036	3357	3357
No. of schools	374	374	170	170	170	170	137	137
No. of federal states	16	16	13	13	13	13	16	16

Note: CS = change score modles. RV = regressor-variable models. 95% confidence intervals in brackets. (c) = Variables are mean centered. Estimates based on 50 multiple-imputed datasets. Federal state fixed effects included (not shown for data protection reasons). Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0), author's calculations.

Table D4: Multilevel models predicting achievement growth in language skills for three student cohorts (non-missing dependent variable)

	Grades 1–3		Grades 5–9		Grades 5–9		Grades 9–12	
	CS	RV	CS	RV	CS	RV	CS	RV
	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate
Social origin (ref: low SES)								
Middle SES (0/1)	0.06*	0.12***	-0.02	0.05	-0.04	0.08	0.05	0.09*
	[0.01,0.11]	[0.07,0.17]	[-0.13,0.09]	[-0.04,0.14]	[-0.15,0.08]	[-0.01,0.18]	[-0.05,0.16]	[0.01,0.17]
High SES (0/1)	0.11***	0.21***	-0.09	0.10*	-0.11	0.17***	0.03	0.15***
	[0.06,0.16]	[0.16,0.26]	[-0.21,0.04]	[0.00,0.20]	[-0.23,0.00]	[0.07,0.27]	[-0.07,0.14]	[0.07,0.23]
Achievement $t-1$ (c)		-0.18***		-0.60***		-0.57***		-0.62***
		[-0.21,-0.16]		[-0.63,-0.57]		[-0.60,-0.54]		[-0.65,-0.60]
Girl (0/1)	-0.03	-0.04*	0.06	0.10**	0.05	0.11***	-0.06	0.08**
	[-0.06,0.00]	[-0.07,-0.01]	[-0.02,0.14]	[0.04,0.17]	[-0.03,0.13]	[0.05,0.18]	[-0.13,0.01]	[0.03,0.14]
Minority student (0/1)	-0.06**	-0.15***	0.10*	-0.07	0.11*	-0.05	0.04	-0.06
	[-0.09,-0.02]	[-0.19,-0.11]	[0.00,0.20]	[-0.15,0.01]	[0.01,0.20]	[-0.13,0.03]	[-0.05,0.12]	[-0.12,0.01]
Age first measure (in months) (c)	-0.01***	-0.01**	0.00	-0.00	0.00	-0.01*	-0.00	-0.01***
	[-0.01,-0.01]	[-0.01,-0.00]	[-0.01,0.01]	[-0.01,0.00]	[-0.00,0.01]	[-0.01,-0.00]	[-0.01,0.00]	[-0.02,-0.01]
Months btw. measurements (c)	0.01*	0.01*	-0.01	0.00	-0.02	0.01	0.02	0.02
	[0.00,0.03]	[0.00,0.03]	[-0.07,0.05]	[-0.05,0.06]	[-0.08,0.04]	[-0.06,0.08]	[-0.03,0.06]	[-0.02,0.06]
School tracks (ref: Academic track)								
Basic req. (0/1)			0.14	-0.78***				
			[-0.03,0.30]	[-0.94,-0.63]				
Basic and extended req. (0/1)			0.08	-0.50***				
			[-0.14,0.30]	[-0.70,-0.30]				
Extended req. (0/1)			0.02	-0.42***				
			[-0.10,0.15]	[-0.54,-0.30]				
Comprehensive (0/1)			0.18	-0.43***				
			[-0.04,0.39]	[-0.62,-0.23]				
Intercept	1.09***	1.07***	1.51***	1.57***	1.58***	1.27***	-0.01	-0.16
	[0.97,1.21]	[0.96,1.18]	[1.22,1.81]	[1.30,1.84]	[1.30,1.87]	[0.94,1.61]	[-0.24,0.22]	[-0.36,0.05]
SD(school)	0.15***	0.14***	0.16***	0.18***	0.16***	0.30***	0.10***	0.13***
	[0.13,0.18]	[0.12,0.16]	[0.11,0.23]	[0.14,0.24]	[0.11,0.24]	[0.24,0.36]	[0.06,0.18]	[0.10,0.17]
SD(student)	0.58***	0.57***	1.05***	0.84***	1.05***	0.84***	0.96**	0.73***
	[0.57,0.59]	[0.56,0.58]	[1.02,1.08]	[0.82,0.86]	[1.02,1.08]	[0.82,0.86]	[0.94,0.99]	[0.71,0.75]
No. of students	5272	5272	2782	2782	2782	2782	3290	3290
No. of schools	353	353	170	170	170	170	137	137
No. of federal states	16	16	13	13	13	13	16	16

Note: CS = change score modles. RV = regressor-variable models. 95% confidence intervals in brackets. (c) = Variables are mean centered. Estimates based on 50 multiple-imputed datasets. Federal state fixed effects included (not shown for data protection reasons). Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0), author's calculations.

D.3 Conditioning on reasoning skills

Some scholars argue that it is necessary to control for intelligence in addition to prior achievement (Murayama et al., 2013; Stienstra et al., 2020), especially if one is interested in SES effects in achievement growth (Marks, 2014). In the paper, I have refrained from doing so because for some cohorts the intelligence measures were not collected until after the first achievement measurement and thus there could be problems with a causal interpretation. In the models in Table D5 and Table D6, I have included reasoning skills as an additional explanatory variable. For information on the intelligence measurements in the NEPS, see Fuß et al., 2016.

Table D5: Multilevel models predicting achievement growth in mathematics for three student cohorts (intelligence as additional control variable)

	Grades 1–4		Grades 5–9		Grades 5–9		Grades 9–12	
	CS	RV	CS	RV	CS	RV	CS	RV
	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate
Social origin (ref: low SES)								
Middle SES (0/1)	0.03	0.21***	0.03	0.08*	0.02	0.11**	0.01	0.06
	[-0.04,0.11]	[0.14,0.28]	[-0.06,0.12]	[0.00,0.16]	[-0.06,0.11]	[0.03,0.19]	[-0.09,0.11]	[-0.03,0.14]
High SES (0/1)	0.08*	0.40***	0.08	0.18***	0.08	0.23***	0.01	0.11*
	[0.00,0.15]	[0.33,0.47]	[-0.01,0.18]	[0.09,0.27]	[-0.01,0.17]	[0.15,0.32]	[-0.09,0.10]	[0.02,0.19]
Achievement $t-1$ (c)		-0.49***		-0.50***		-0.46***		-0.47***
		[-0.52,-0.47]		[-0.54,-0.47]		[-0.50,-0.43]		[-0.50,-0.45]
Girl (0/1)	0.09***	-0.05*	-0.03	-0.17***	-0.03	-0.16***	-0.03	-0.29***
	[0.04,0.14]	[-0.09,-0.00]	[-0.09,0.04]	[-0.23,-0.12]	[-0.09,0.03]	[-0.21,-0.10]	[-0.09,0.04]	[-0.35,-0.23]
Minority student (0/1)	0.11***	-0.04	0.02	-0.09*	0.02	-0.08*	0.02	-0.08*
	[0.05,0.17]	[-0.09,0.02]	[-0.06,0.10]	[-0.16,-0.02]	[-0.06,0.10]	[-0.15,-0.01]	[-0.05,0.10]	[-0.15,-0.02]
Age first measure (in months) (c)	-0.03***	-0.02***	-0.00	-0.01**	-0.00	-0.01***	-0.01**	-0.01***
	[-0.03,-0.02]	[-0.03,-0.02]	[-0.01,0.00]	[-0.01,-0.00]	[-0.01,0.00]	[-0.02,-0.01]	[-0.01,-0.00]	[-0.02,-0.01]
Months btw. measurements (c)	0.10***	0.04***	-0.01	-0.02	-0.01	0.02	-0.01	-0.01
	[0.08,0.13]	[0.02,0.06]	[-0.08,0.06]	[-0.08,0.04]	[-0.07,0.06]	[-0.06,0.09]	[-0.05,0.03]	[-0.04,0.03]
Reasoning test score	0.00	0.07***	-0.02*	0.06***	-0.02**	0.06***	-0.03**	0.06***
	[-0.01,0.01]	[0.06,0.08]	[-0.03,-0.00]	[0.04,0.07]	[-0.03,-0.01]	[0.05,0.08]	[-0.05,-0.01]	[0.04,0.08]
School tracks (ref: Academic track)								
Basic req. (0/1)			0.10	-0.54***				
			[-0.05,0.25]	[-0.69,-0.40]				
Basic and extended req. (0/1)			0.03	-0.40***				
			[-0.16,0.21]	[-0.58,-0.23]				
Extended req. (0/1)			-0.09	-0.43***				
			[-0.20,0.02]	[-0.53,-0.32]				
Comprehensive (0/1)			-0.01	-0.47***				
			[-0.19,0.18]	[-0.65,-0.29]				
Intercept	2.52***	2.05***	1.62***	1.26***	1.62***	0.97***	1.02***	0.03
	[2.33,2.72]	[1.90,2.21]	[1.36,1.87]	[1.02,1.50]	[1.38,1.87]	[0.70,1.24]	[0.72,1.31]	[-0.23,0.28]
SD(school)	0.24***	0.17***	0.17***	0.17***	0.18***	0.24***	0.11***	0.09***
	[0.21,0.28]	[0.14,0.21]	[0.13,0.22]	[0.14,0.21]	[0.14,0.23]	[0.20,0.29]	[0.07,0.17]	[0.05,0.15]
SD(student)	0.95***	0.84***	0.87***	0.77***	0.87***	0.78***	0.90***	0.77***
	[0.93,0.97]	[0.82,0.85]	[0.85,0.89]	[0.75,0.79]	[0.85,0.89]	[0.75,0.80]	[0.88,0.93]	[0.75,0.79]
No. of students	6866	6866	3461	3461	3461	3461	3946	3946
No. of schools	374	374	171	171	171	171	137	137
No. of federal states	16	16	13	13	13	13	16	16

Note: CS = change score modles. RV = regressor-variable models. 95% confidence intervals in brackets. (c) = Variables are mean centered. Estimates based on 50 multiple-imputed datasets. Federal state fixed effects included (not shown for data protection reasons). Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0), author's calculations.

Table D6: Multilevel models predicting achievement growth in language skills for three student cohorts (intelligence as additional control variable)

	Grades 1–3		Grades 5–9		Grades 5–9		Grades 9–12	
	CS	RV	CS	RV	CS	RV	CS	RV
	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate
Social origin (ref: low SES)								
Middle SES (0/1)	0.04	0.10***	-0.02	0.05	-0.03	0.08	0.06	0.09*
	[-0.00,0.09]	[0.06,0.15]	[-0.13,0.09]	[-0.04,0.13]	[-0.13,0.08]	[-0.01,0.17]	[-0.04,0.17]	[0.01,0.17]
High SES (0/1)	0.09***	0.19***	-0.09	0.09	-0.10	0.16***	0.04	0.15***
	[0.04,0.14]	[0.14,0.24]	[-0.20,0.03]	[-0.00,0.19]	[-0.21,0.02]	[0.07,0.26]	[-0.06,0.14]	[0.07,0.23]
Achievement $t-1$ (c)		-0.19***		-0.62***		-0.59***		-0.63***
		[-0.22,-0.17]		[-0.65,-0.59]		[-0.62,-0.56]		[-0.66,-0.61]
Girl (0/1)	-0.03	-0.04**	0.04	0.12***	0.04	0.13***	-0.07*	0.10***
	[-0.06,0.00]	[-0.08,-0.01]	[-0.04,0.12]	[0.06,0.18]	[-0.04,0.12]	[0.06,0.19]	[-0.14,-0.01]	[0.04,0.15]
Minority student (0/1)	-0.05**	-0.15***	0.07	-0.07	0.07	-0.06	0.03	-0.05
	[-0.09,-0.01]	[-0.19,-0.11]	[-0.03,0.16]	[-0.15,0.01]	[-0.02,0.17]	[-0.14,0.02]	[-0.06,0.11]	[-0.11,0.01]
Age first measure (in months) (c)	-0.01***	-0.01**	0.00	-0.00	0.00	-0.01	-0.01*	-0.01***
	[-0.01,-0.00]	[-0.01,-0.00]	[-0.01,0.01]	[-0.01,0.00]	[-0.00,0.01]	[-0.01,0.00]	[-0.01,-0.00]	[-0.02,-0.01]
Months btw. measurements (c)	0.02**	0.01**	-0.00	0.00	-0.00	0.01	0.02	0.01
	[0.01,0.03]	[0.00,0.02]	[-0.06,0.06]	[-0.05,0.05]	[-0.06,0.06]	[-0.06,0.07]	[-0.02,0.07]	[-0.02,0.05]
Reasoning test score	0.01*	0.02***	-0.03***	0.04***	-0.03***	0.05***	-0.06***	0.03***
	[0.00,0.01]	[0.01,0.03]	[-0.05,-0.02]	[0.03,0.06]	[-0.05,-0.02]	[0.04,0.07]	[-0.08,-0.04]	[0.02,0.05]
School tracks (ref: Academic track)								
Basic req. (0/1)			0.05	-0.71***				
			[-0.12,0.22]	[-0.86,-0.56]				
Basic and extended req. (0/1)			0.05	-0.44***				
			[-0.16,0.26]	[-0.63,-0.26]				
Extended req. (0/1)			-0.01	-0.38***				
			[-0.13,0.11]	[-0.49,-0.27]				
Comprehensive (0/1)			0.17	-0.36***				
			[-0.03,0.37]	[-0.54,-0.18]				
Intercept	1.06***	0.97***	1.78***	1.22***	1.83***	0.93***	0.54***	-0.47***
	[0.95,1.17]	[0.86,1.08]	[1.48,2.07]	[0.97,1.48]	[1.56,2.11]	[0.64,1.22]	[0.24,0.84]	[-0.72,-0.22]
SD(school)	0.12***	0.11***	0.15***	0.15***	0.15***	0.25***	0.10***	0.11***
	[0.10,0.15]	[0.09,0.14]	[0.10,0.22]	[0.11,0.21]	[0.10,0.22]	[0.20,0.31]	[0.05,0.17]	[0.08,0.16]
SD(student)	0.59***	0.57***	1.04**	0.84***	1.04**	0.84***	0.96**	0.73***
	[0.57,0.60]	[0.56,0.58]	[1.02,1.07]	[0.81,0.86]	[1.02,1.07]	[0.82,0.86]	[0.94,0.99]	[0.71,0.75]
No. of students	6865	6865	3350	3350	3350	3350	3908	3908
No. of schools	374	374	172	172	172	172	137	137
No. of federal states	16	16	13	13	13	13	16	16

Note: CS = change score modles. RV = regressor-variable models. 95% confidence intervals in brackets. (c) = Variables are mean centered. Estimates based on 50 multiple-imputed datasets. Federal state fixed effects included (not shown for data protection reasons). Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0), author's calculations.

D.4 Measurement error

The negative association between prior achievement and subsequent growth could be driven by measurement error in prior achievement, which could subsequently lead to a “regression to the mean” and thus bias SES group effects as well as the effects of prior achievement on subsequent achievement growth (Allison, 1990; Kelly and Ye, 2017). This problem can be addressed in several ways. Since we are concerned with biased results due to measurement error in prior achievement, only regressor variable approach models (which control for prior achievement) are presented.

First, I used the test score rank (divided by number of students) instead of the absolute test score of the first measurement (Betz, 2013). This is an adequate procedure if one can assume that the measurement error distorts the absolute test scores but not the ranking within students. These models are shown in Tables D7 and D8.

Second, I used the test score rank of the first measurement (divided by 1,000) as an instrument for the absolute test score of the first measurement (Betz, 2013; Feinstein, 2003). Unfortunately, there is no achievement score in the data that would be a better instrument such as a mathematics or reading achievement measurement from a measurement a few days earlier. Therefore, only the achievement rank is available to me as an — admittedly weak — instrument. In a first step, I regressed the absolute achievement on the achievement rank at the first measurement time (see Equation 1). Then I predicted the absolute achievement for each student (i) (based on the regression results) at the first measurement time (see Equation 2). In a second step, I included this predicted absolute achievement as a predictor in my multilevel models (see Equation 3). The results of the second-step regressions are shown in Tables D9 and D10.

$$Y_{t-1} \quad = \quad \gamma_0 + \gamma_1 Y_{t-1R} + r \tag{1}$$

where

Y_{t-1} is the achievement score at the first measurement;

Y_{t-1R} is the rank of the achievement score at the first measurement;

γ are the regression coefficients;

r is the error term.

$$Y_{t-1_{pred_i}} = \gamma_0 + \gamma_1 Y_{t-1R_i} \tag{2}$$

where

$Y_{t-1_{pred_i}}$ is the predicted achievement score at the first measurement for student i based on regression results from Eq. 1; γ are the regression coefficients;

Y_{t-1R} is the rank of the achievement score at the first measurement for student i.

$$G_{ij} = Y_{t_{ij}} - Y_{t-1_{ij}} \quad = \quad \gamma_{00} + u_{0j} + \gamma_{10} Y_{t-1_{predij}} + \gamma_{X0} \mathbf{X}_{ij} + + \gamma_{Z0} \mathbf{Z}_{ij} + r_{ij} \tag{3}$$

where

G_{ij} is the achievement growth btw. at first and last measurement for student i in school j;

$Y_{t_{ij}}$ is the achievement score at the last measurement for student i in school j;

$Y_{t-1_{ij}}$ is the achievement score at the first measurement for student i in school j;

γ are the regression coefficients;

u_{0j} ist the random intercept on the school level;

$Y_{t-1_{predij}}$ is the predicted achievement score at the first measurement for student i in school j;

$\mathbf{X}$ is a vector for all the individual-level model variables;

$\mathbf{Z}$ is a vector for the federal states dummies;

r is the error term for student i in school j.

Third, following the approach of Jerrim and Vignoles (2013), I used the mean of the test scores from the first two measurements to predict growth between the second and third measurements (only for student cohorts with three measuremnt points). These models are shown in Tables D11 and D12.

Table D7: Multilevel models predicting achievement growth in mathematics for three student cohorts (achievement rank as predictor instead of absolute achievement)

	Grades 1–4	Grades 5–9		Grades 9–12
	Estimate	Estimate	Estimate	Estimate
Social origin (ref: low SES)				
Middle SES (0/1)	0.22***	0.07	0.10*	0.05
	[0.15,0.29]	[-0.01,0.15]	[0.02,0.18]	[-0.04,0.14]
High SES (0/1)	0.40***	0.17***	0.22***	0.09
	[0.33,0.47]	[0.08,0.26]	[0.14,0.31]	[-0.00,0.18]
Achievement rank $t-1$ (c)	-0.44***	-0.46***	-0.41***	-0.46***
	[-0.47,-0.41]	[-0.50,-0.42]	[-0.44,-0.37]	[-0.49,-0.43]
Girl (0/1)	-0.00	-0.15***	-0.13***	-0.27***
	[-0.05,0.05]	[-0.21,-0.10]	[-0.19,-0.08]	[-0.33,-0.21]
Minority student (0/1)	-0.01	-0.08*	-0.07	-0.08*
	[-0.07,0.04]	[-0.15,-0.01]	[-0.14,0.00]	[-0.15,-0.01]
Age first measure (in months) (c)	-0.02***	-0.01*	-0.01***	-0.01***
	[-0.03,-0.02]	[-0.01,-0.00]	[-0.02,-0.00]	[-0.02,-0.01]
Months btw. measurements (c)	0.05***	-0.02	0.02	-0.01
	[0.03,0.08]	[-0.08,0.05]	[-0.06,0.09]	[-0.05,0.02]
School tracks (ref: Academic track)				
Basic req. (0/1)		-0.54***		
		[-0.68,-0.39]		
Basic and extended req. (0/1)		-0.42***		
		[-0.60,-0.24]		
Extended req. (0/1)		-0.44***		
		[-0.56,-0.33]		
Comprehensive (0/1)		-0.47***		
		[-0.65,-0.29]		
Intercept	2.48***	1.70***	1.44***	0.66***
	[2.33,2.64]	[1.48,1.92]	[1.18,1.70]	[0.46,0.85]
SD(school)	0.18***	0.17***	0.25***	0.09***
	[0.15,0.22]	[0.14,0.22]	[0.20,0.30]	[0.05,0.15]
SD(school)	0.87***	0.80***	0.80***	0.80***
	[0.85,0.89]	[0.78,0.82]	[0.78,0.82]	[0.78,0.82]
No. of students	6866	3461	3461	3946
No. of schools	374	171	171	137
No. of federal states	16	13	13	16

Note: Regressor-variable models only. 95% confidence intervals in brackets. (c) = Variables are mean centered. Estimates based on 50 multiple-imputed datasets. Federal state fixed effects included (not shown for data protection reasons). Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0), author's calculations.

Table D8: Multilevel models predicting achievement growth in language skills for three student cohorts (achievement rank as predictor instead of absolute achievement)

	Grades 1–3	Grades 5–9		Grades 9–12
	Estimate	Estimate	Estimate	Estimate
Social origin (ref: low SES)				
Middle SES (0/1)	0.10***	0.05	0.08	0.09*
	[0.06,0.14]	[-0.04,0.14]	[-0.01,0.17]	[0.00,0.17]
High SES (0/1)	0.19***	0.09	0.16**	0.15***
	[0.14,0.24]	[-0.01,0.18]	[0.06,0.25]	[0.07,0.23]
Achievement rank $t-1$ (c)	-0.13***	-0.71***	-0.66***	-0.60***
	[-0.15,-0.11]	[-0.75,-0.67]	[-0.70,-0.62]	[-0.63,-0.58]
Girl (0/1)	-0.04*	0.09**	0.10**	0.06*
	[-0.07,-0.00]	[0.03,0.16]	[0.04,0.17]	[0.01,0.12]
Minority student (0/1)	-0.13***	-0.07	-0.05	-0.04
	[-0.17,-0.09]	[-0.15,0.02]	[-0.13,0.03]	[-0.11,0.02]
Age first measure (in months) (c)	-0.01**	-0.00	-0.01	-0.01***
	[-0.01,-0.00]	[-0.01,0.00]	[-0.01,0.00]	[-0.02,-0.01]
Months btw. measurements (c)	0.02**	0.00	0.01	0.02
	[0.00,0.03]	[-0.05,0.06]	[-0.06,0.08]	[-0.02,0.06]
School tracks (ref: Academic track)				
Basic req. (0/1)		-0.76***		
		[-0.91,-0.61]		
Basic and extended req. (0/1)		-0.50***		
		[-0.69,-0.31]		
Extended req. (0/1)		-0.43***		
		[-0.54,-0.32]		
Comprehensive (0/1)		-0.39***		
		[-0.58,-0.21]		
Intercept	1.08***	1.59***	1.35***	-0.17
	[0.98,1.19]	[1.36,1.83]	[1.05,1.64]	[-0.36,0.03]
SD(school)	0.11***	0.15***	0.28***	0.11***
	[0.09,0.14]	[0.11,0.21]	[0.22,0.34]	[0.08,0.16]
SD(student)	0.58***	0.86***	0.86***	0.76***
	[0.56,0.59]	[0.84,0.89]	[0.84,0.89]	[0.74,0.78]
No. of students	6865	3350	3350	3908
No. of schools	374	172	172	137
No. of federal states	16	13	13	16

Note: Regressor-variable models only. 95% confidence intervals in brackets. (c) = Variables are mean centered. Estimates based on 50 multiple-imputed datasets. Federal state fixed effects included (not shown for data protection reasons). Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0), author's calculations.

Table D9: Multilevel models predicting achievement growth in mathematics for three student cohorts (predicted achievement as predictor instead of absolute achievement)

	Grades 1–4	Grades 5–9		Grades 9–12
	Estimate	Estimate	Estimate	Estimate
Equation 1: DV: Achievement $t-1$				
Achievement rank $t-1$ (c)	0.57	1.16	1.16	0.98
	[0.57,0.58]	[1.15,1.17]	[1.15,1.17]	[0.97,0.99]
Intercept	1.68	0.10	0.10	1.03
	[1.67,1.68]	[0.09,0.11]	[0.09,0.11]	[1.02,1.04]
Equation 2: DV: Achievement growth				
Social origin (ref: low SES)				
Middle SES (0/1)	0.22***	0.07	0.10*	0.05
	[0.15,0.29]	[-0.01,0.15]	[0.02,0.18]	[-0.04,0.14]
High SES (0/1)	0.40***	0.17***	0.22***	0.09
	[0.33,0.47]	[0.08,0.26]	[0.14,0.31]	[-0.00,0.18]
IV (predicted achievement $t-1$)	-0.41***	-0.42***	-0.37***	-0.42***
	[-0.43,-0.38]	[-0.45,-0.38]	[-0.40,-0.33]	[-0.45,-0.40]
Girl (0/1)	-0.00	-0.16***	-0.14***	-0.27***
	[-0.05,0.05]	[-0.21,-0.10]	[-0.19,-0.08]	[-0.33,-0.21]
Minority student (0/1)	-0.01	-0.08*	-0.07	-0.08*
	[-0.07,0.04]	[-0.15,-0.01]	[-0.14,0.00]	[-0.15,-0.01]
Age first measure (in months) (c)	-0.02***	-0.01*	-0.01***	-0.01***
	[-0.03,-0.02]	[-0.01,-0.00]	[-0.02,-0.00]	[-0.02,-0.01]
Months btw. measurements (c)	0.05***	-0.02	0.02	-0.01
	[0.03,0.08]	[-0.08,0.05]	[-0.06,0.09]	[-0.05,0.02]
School tracks (ref: Academic track)				
Basic req. (0/1)		-0.54***		
		[-0.68,-0.39]		
Basic and extended req. (0/1)		-0.42***		
		[-0.60,-0.24]		
Extended req. (0/1)		-0.44***		
		[-0.55,-0.33]		
Comprehensive (0/1)		-0.47***		
		[-0.65,-0.29]		
Intercept	3.17***	1.74***	1.48***	1.09***
	[3.01,3.33]	[1.52,1.97]	[1.22,1.74]	[0.90,1.29]
SD(school)	0.18***	0.17***	0.25***	0.09***
	[0.15,0.22]	[0.14,0.22]	[0.20,0.30]	[0.05,0.15]
SD(student)	0.87***	0.80***	0.80***	0.80***
	[0.85,0.89]	[0.78,0.82]	[0.78,0.82]	[0.78,0.82]
No. of students	6866	3461	3461	3946
No. of schools	374	171	171	137
No. of federal states	16	13	13	16

Note: Regressor-variable models only. 95% confidence intervals in brackets. (c) = Variables are mean centered. Estimates based on 50 multiple-imputed datasets. Federal state fixed effects included (not shown for data protection reasons). Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0), author's calculations.

Table D10: Multilevel models predicting achievement growth in language skills for three student cohorts (predicted achievement as predictor instead of absolute achievement)

	Grades 1–4	Grades 5–9		Grades 9–12
	Estimate	Estimate	Estimate	Estimate
Equation 1: DV: Achievement $t-1$				
Achievement rank $t-1$ (c)	0.42	1.30	1.30	0.92
	[0.42,0.43]	[1.29,1.31]	[1.29,1.31]	[0.91,0.93]
Intercept	1.42	0.11	0.11	0.87
	[1.41,1.42]	[0.10,0.12]	[0.10,0.12]	[0.86,0.87]
Equation 2: DV: Achievement growth				
Social origin (ref: low SES)				
Middle SES (0/1)	0.10***	0.05	0.08	0.09*
	[0.05,0.14]	[-0.04,0.14]	[-0.01,0.17]	[0.00,0.17]
High SES (0/1)	0.19***	0.09	0.16**	0.15***
	[0.14,0.24]	[-0.01,0.18]	[0.06,0.25]	[0.07,0.23]
IV (predicted achievement $t-1$)	-0.16***	-0.59***	-0.55***	-0.60***
	[-0.19,-0.14]	[-0.62,-0.56]	[-0.58,-0.52]	[-0.63,-0.58]
Girl (0/1)	-0.04*	0.09**	0.10**	0.06*
	[-0.07,-0.00]	[0.03,0.16]	[0.04,0.17]	[0.01,0.12]
Minority student (0/1)	-0.13***	-0.07	-0.05	-0.04
	[-0.17,-0.09]	[-0.15,0.02]	[-0.13,0.03]	[-0.11,0.02]
Age first measure (in months) (c)	-0.01**	-0.00	-0.01	-0.01***
	[-0.01,-0.00]	[-0.01,0.00]	[-0.01,0.00]	[-0.02,-0.01]
Months btw. measurements (c)	0.02**	0.00	0.01	0.02
	[0.00,0.03]	[-0.05,0.06]	[-0.06,0.08]	[-0.02,0.06]
School tracks (ref: Academic track)				
Basic req. (0/1)		-0.76***		
		[-0.91,-0.61]		
Basic and extended req. (0/1)		-0.50***		
		[-0.69,-0.31]		
Extended req. (0/1)		-0.43***		
		[-0.54,-0.32]		
Comprehensive (0/1)		-0.39***		
		[-0.57,-0.21]		
Intercept	1.32***	1.66***	1.41***	0.36***
	[1.21,1.43]	[1.42,1.90]	[1.11,1.70]	[0.16,0.56]
SD(school)	0.11***	0.15***	0.28***	0.11***
	[0.09,0.14]	[0.11,0.21]	[0.22,0.34]	[0.08,0.16]
SD(student)	0.58***	0.86***	0.86***	0.76***
	[0.56,0.59]	[0.84,0.89]	[0.84,0.89]	[0.74,0.78]
No. of students	6865	3350	3350	3908
No. of schools	374	172	172	137
No. of federal states	16	13	13	16

Note: Regressor-variable models only. 95% confidence intervals in brackets. (c) = Variables are mean centered. Estimates based on 50 multiple-imputed datasets. Federal state fixed effects included (not shown for data protection reasons). Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0), author's calculations.

Table D11: Multilevel models predicting achievement growth in mathematics for three student cohorts (mean achievement as predictor)

	Grades 2–4	Grades 7–9	
	Estimate	Estimate	Estimate
Social origin (ref: low SES)			
Middle SES (0/1)	0.17***	0.03	0.05
	[0.10,0.25]	[-0.05,0.12]	[-0.04,0.13]
High SES (0/1)	0.30***	0.06	0.10*
	[0.22,0.37]	[-0.03,0.15]	[0.01,0.19]
Mean achievement first two measures	-0.31***	-0.32***	-0.28***
	[-0.34,-0.29]	[-0.36,-0.29]	[-0.31,-0.25]
Girl (0/1)	0.11***	-0.06	-0.04
	[0.06,0.16]	[-0.12,0.00]	[-0.10,0.02]
Minority student (0/1)	-0.02	-0.10**	-0.09*
	[-0.07,0.04]	[-0.18,-0.03]	[-0.16,-0.02]
Months btw. measurements (c)	0.01	0.03	0.03
	[-0.03,0.04]	[-0.02,0.07]	[-0.01,0.08]
Age second measure (in months) (c)	-0.02***	-0.01*	-0.01**
	[-0.03,-0.01]	[-0.01,-0.00]	[-0.01,-0.00]
School tracks (ref: Academic track)			
Basic req. (0/1)		-0.31***	
		[-0.45,-0.16]	
Basic and extended req. (0/1)		-0.31***	
		[-0.49,-0.14]	
Extended req. (0/1)		-0.30***	
		[-0.41,-0.19]	
Comprehensive (0/1)		-0.22*	
		[-0.39,-0.05]	
Intercept	2.73***	0.96***	0.78***
	[2.56,2.89]	[0.75,1.18]	[0.56,1.00]
SD(school)	0.17***	0.15***	0.18***
	[0.14,0.21]	[0.11,0.20]	[0.14,0.23]
SD(student)	0.87***	0.82***	0.82***
	[0.85,0.89]	[0.80,0.84]	[0.80,0.85]
No. of students	6866	3463	3463
No. of schools	374	171	171
No. of federal states	16	13	13

Note: Regressor-variable models only. 95% confidence intervals in brackets. (c) = Variables are mean centered. Estimates based on 50 multiple-imputed datasets. Federal state fixed effects included (not shown for data protection reasons). Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0), author's calculations.

Table D12: Multilevel models predicting achievement growth in language skills for three student cohorts (mean acheievment as predictor)

	Grades 7–9	
	Estimate	Estimate
Social origin (ref: low SES)		
Middle SES (0/1)	0.05	0.09
	[-0.05,0.14]	[-0.01,0.18]
High SES (0/1)	0.08	0.15**
	[-0.02,0.18]	[0.05,0.25]
Mean achievement first two measures	-0.50***	-0.45***
	[-0.54,-0.46]	[-0.48,-0.41]
Girl (0/1)	-0.01	-0.01
	[-0.08,0.06]	[-0.08,0.06]
Minority student (0/1)	-0.12**	-0.11*
	[-0.21,-0.04]	[-0.19,-0.02]
Age second measure (in months) (c)	0.00	-0.00
	[-0.01,0.01]	[-0.01,0.00]
Months btw. measurements (c)	0.01	0.02
	[-0.03,0.06]	[-0.03,0.06]
School tracks (ref: Academic track)		
Basic req. (0/1)	-0.52***	
	[-0.67,-0.37]	
Basic and extended req. (0/1)	-0.38***	
	[-0.57,-0.20]	
Extended req. (0/1)	-0.25***	
	[-0.35,-0.14]	
Comprehensive (0/1)	-0.31***	
	[-0.48,-0.14]	
Intercept	0.97***	0.76***
	[0.74,1.21]	[0.53,1.00]
SD(school)	0.11***	0.15***
	[0.06,0.19]	[0.10,0.22]
SD(student)	0.95***	0.95***
	[0.92,0.97]	[0.93,0.98]
No. of students	3350	3350
No. of schools	172	172
No. of federal states	13	13

Note: Regressor-variable models only. 95% confidence intervals in brackets. (c) = Variables are mean centered. Estimates based on 50 multiple-imputed datasets. Federal state fixed effects included (not shown for data protection reasons). Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0), author's calculations.

D.5 Growth curve models

To estimate linear growth using growth curve models, at least three measurement time points are needed (King et al., 2018; Raudenbush and Bryk, 2003). Since three measurement time points are not available for all three cohorts, I have refrained from estimating growth curve models in the paper. For cohorts SC2 (only mathematics) and SC3, for which three achievement test score measurements are available, I was able to estimate growth curve models. I estimated these growth curve models in a fixed effect framework (Allison, 2009).

I have estimated fixed effect growth curves because these models keep all unobserved individual factors fixed. Furthermore, with these models it is possible to estimate group differences with a limited number of additional parameters. Since I am interested in differences in time-constant group variables (high vs. low performing students at first measurement point and high vs. low SES), I estimated rather simple models (see Table D13).

I used three SES groups (low, middle, high), the first achievement measure, the time variable (time since first measurement in years) and interactions between time and SES and time and achievement is included. Model 1 (M1) contains the interaction between time and SES. Model 2 (M2) contains the interaction between time and prior achievement. Model 3 (M3) contains both interactions to account for the possible path-dependent cumulative advantage. Based on Model 3, I calculated the predicted effects (Klein, 2014) of SES and prior achievement over the entire observation period in order to compare these effects with the effects presented in the paper.

In general, these models confirm the results presented in the paper: SES gaps remain largely constant, once controlled for prior achievement SES gaps grow, and formerly lower performing students are able to close the gap slightly to high performing students.

Table D13: Fixed effect growth curve models predicting achievement growth

	SC2 Primary school Grades 1–4 (2.68 years)			SC3 Lower secondary school Grades 5–9 (4.01 years)			SC3 Lower secondary school Grades 5–9 (4.44 years)		
	Mathematics			Mathematics			Language skills		
	M1	M2	M3	M1	M2	M3	M1	M2	M3
Time since first measurement (in years)	1.08***	1.30***	1.23***	0.37***	0.39***	0.33***	0.29***	0.28***	0.24***
	[1.06,1.10]	[1.29,1.32]	[1.20,1.25]	[0.36,0.39]	[0.38,0.39]	[0.32,0.35]	[0.28,0.31]	[0.28,0.29]	[0.23,0.26]
Middle SES x time	0.02		0.09***	0.00		0.05***	-0.02		0.04**
	[-0.01,0.04]		[0.07,0.12]	[-0.02,0.02]		[0.03,0.07]	[-0.04,0.00]		[0.01,0.06]
Middle SES x time r	0.04**		0.17***	0.02		0.10***	-0.04***		0.07***
	[0.02,0.07]		[0.14,0.20]	[-0.00,0.04]		[0.08,0.12]	[-0.06,-0.02]		[0.05,0.10]
Achievement first measure x time		-0.12***	-0.14***		-0.07***	-0.08***		-0.10***	-0.11***
		[-0.13,-0.11]	[-0.14,-0.13]		[-0.07,-0.06]	[-0.08,-0.07]		[-0.11,-0.10]	[-0.12,-0.10]
Intercept	1.61***	1.61***	1.61***	0.11***	0.11***	0.11***	0.17***	0.17***	0.17***
	[1.59,1.62]	[1.59,1.62]	[1.59,1.62]	[0.09,0.13]	[0.09,0.13]	[0.09,0.13]	[0.15,0.20]	[0.15,0.20]	[0.15,0.20]
No. of students	6,866	6,866	6,866	3,463	3,463	3,463	3,350	3,350	3,350
Rho	0.68	0.75	0.75	0.75	0.80	0.80	0.66	0.76	0.76
σ_u	1.01	1.15	1.14	1.08	1.22	1.22	1.09	1.31	1.31
σ_e	0.68	0.66	0.65	0.63	0.61	0.60	0.78	0.73	0.73
Predicted effects over obs. period[1]									
High vs. low SES			0.46			0.39			0.33
			[0.39,0.53]			[0.31,0.48]			[0.22,0.43]
Achievement gap in $t-1$ (0 vs. 1)			-0.36			-0.31			-0.49
			[-0.39,-0.34]			[-0.34,-0.28]			[-0.52,-0.46]

Note: 95% confidence interval in brackets. Estimates based on 50 multiple-imputed datasets. [1] Predicted multiplying interaction effect with years between first and last measuremnt Source: NEPS SC2 (9.0.0), SC3 (11.0.1), SC4(12.0.0), author's calculations.

E Data sets and replication material

The datasets used in this study (without federal state information) can be downloaded after signing a data usage agreement.

NEPS SC2: dx.doi.org/10.5157/NEPS:SC2:9.0.0

NEPS SC3: dx.doi.org/10.5157/NEPS:SC3:11.0.1

NEPS SC4: dx.doi.org/10.5157/NEPS:SC4:12.0.0

The Stata-code for this article is available in the Additional Online Material.

If you want to replicate my results, you need access to the RemoteNEPS environment. All analyses were run in the RemoteNEPS environment. Before you can access this environment you need to sign a data usage contract and the RemoteNEPS Supplemental Agreement. https://www.neps-data.de/Data-Center/Data-Access/RemoteNEPS

References

Allison, P. D. (1990). Change scores as dependent variables in regression analysis, *Sociological methodology*, **20**, 93–114.

Allison, P. D. (2009). *Fixed Effects Regression Models.* Los Angeles: Sage Publications.

Betz, T. (2013). Robust Estimation with Nonrandom Measurement Error and Weak Instruments, *Political Analysis*, **21**, 86–96.

Bol, T., Witschge, J., Van de Werfhorst, H. and Dronkers, J. (2014). Curricular Tracking and Central Examinations: Counterbalancing the Impact of Social Background on Student Achievement in 36 Countries, *Social Forces*, **92**, 1545–1572.

Dollmann, J. (2016). Less Choice, Less Inequality? A Natural Experiment on Social and Ethnic Differences in Educational Decision-Making, *European Sociological Review*, **32**, 203–215.

Duchhardt, C. and Gerdes, A. (2012). NEPS Technical Report for Mathematics: Scaling results of Starting Cohort 3 in fifth grade, *NEPS Working Paper No. 19.*

Duchhardt, C. and Gerdes, A. (2013). NEPS Technical Report for Mathematics: Scaling results of Starting Cohort 4 in ninth grade, *NEPS Working Paper No. 22.*

Esser, H. and Relikowski, I. (2015). Is Ability Tracking (Really) Responsible for Educational Inequalities in Achievement? A Comparison between the Country States Bavaria and Hesse in Germany, *IZA Discussion Paper.*

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

Fischer, L. and Durda, T. (2020). NEPS Technical Report for Receptive Vocabulary: Scaling Results of Starting Cohort 2 for Kindergarten (Wave 1), Grade 1 (Wave 3) and Grade 3 (Wave 5), *NEPS Survey Paper No. 65.*

Fischer, L., Rohm, T. and Gnambs, T. (2017). NEPS Technical Report for Mathematics: Scaling Results of Starting Cohort 4 for Grade 12, *NEPS Survey Paper No. 12.*

Fischer, L., Rohm, T., Gnambs, T. and Carstensen, C. H. (2016). Linking the Data of the Competence Tests, *NEPS Survey Paper No 1.*

Fuß, D., Gnambs, T., Lockl, K. and Attig, M. (2016). Competence data in NEPS: Overview of measures and variable naming conventions (Starting Cohorts 1 to 6), *Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study.*

Gnambs, T., Fischer, L. and Rohm, T. (2017). NEPS Technical Report for Reading: Scaling Results of Starting Cohort 4 for Grade 12, *NEPS Survey Paper No. 13*.

Haberkorn, K., Pohl, S., Hardt, K. and Wiegand, E. (2012). NEPS Technical Report for Reading: Scaling Results of Starting Cohort 4 in Ninth Grade, *NEPS Survey Paper No. 16*.

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**, 887–906.

Kelly, S. and Ye, F. (2017). Accounting for the Relationship Between Initial Status and Growth in Regression Models, *The Journal of Experimental Education*, **85**, 353–375.

King, K. M., Littlefield, A. K., McCabe, C. J., Mills, K. L., Flournoy, J. and Chassin, L. (2018). Longitudinal modeling in developmental neuroimaging research: Common challenges, and solutions from developmental psychology, *Developmental Cognitive Neuroscience*, **33**, 54–72.

Klein, D. (2014). *MIMRGNS: Stata module to run margins after mi estimate*. Statistical Software Components, Boston College Department of Economics.

Krannich, M., Jost, O., Rohm, T., Koller, I., Pohl, S., Haberkorn, K., Carstensen, C. H., Fischer, L. and Gnambs, T. (2017). NEPS Technical Report for Reading: Scaling Results of Starting Cohort 3 for Grade 7, *Update NEPS Survey Paper No. 14*.

Marks, G. N. (2014). *Education, Social Background and Cognitive Ability. The decline of the social.* New York: Routledge.

Murayama, K., Pekrun, R., Lichtenfeld, S. and vom Hofe, R. (2013). Predicting Long-Term Growth in Students' Mathematics Achievement: The Unique Contributions of Motivation and Cognitive Strategies, *Child Development*, **84**, 1475–1490.

Neugebauer, M., Reimer, D., Schindler, S. and Stocké, V. (2013). Inequality in Transitions to Secondary School and Tertiary Education in Germany. In Jackson, M. (Eds.), *Determined to succeed? Performance versus Choice in Educational Attainment*, 56–88. Stanford: Stanford University Press.

Pohl, S. and Carstensen, C. H. (2012). NEPS Technical Report-Scaling the Data of the Competence Tests, *NEPS Working Papers No. 14*.

Pohl, S., Haberkorn, K., Hardt, K. and Wiegand, E. (2012). NEPS Technical Report for Reading: Scaling Results of Starting Cohort 3 in Fifth Grade, *NEPS Survey Paper No. 15*.

Raudenbush, S. W. and Bryk, A. S. (2003). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks: Sage Publications.

Scharl, A., Fischer, L., Gnambs, T. and Rohm, T. (2017). NEPS Technical Report for Reading: Scaling Results of Starting Cohort 3 for Grade 9, *NEPS Survey Paper No. 20*.

Schneider, T. (2008). Social Inequality in Educational Participation in the German School System in a Longitudinal Perspective: Pathways into and out of the most Prestigious School Track, *European Sociological Review*, **24**, 511–526.

Schnittjer, I. and Fischer, L. (2018). NEPS Technical Report for Mathematics: Scaling Results of Starting Cohort 2 for Grade 1, *NEPS Survey Paper No. 46*.

Schnittjer, I. and Gerken, A.-L. (2017). NEPS Technical Report for Mathematics: Scaling Results of Starting Cohort 3 in Grade 7, *NEPS Survey Paper No. 16*.

Schnittjer, I. and Gerken, A.-L. (2018). NEPS Technical Report for Mathematics: Scaling Results of Starting Cohort 2 for Grade 2, *NEPS Survey Paper No. 47*.

Schnittjer, I., Gerken, A.-L. and Petersen, L. A. (2020). NEPS Technical Report for Mathematics: Scaling Results of Starting Cohort 2 in Grade 4, *NEPS Survey Paper No. 69*.

Stienstra, K., Maas, I., Knigge, A. and Schulz, W. (2020). Resource Compensation or Multiplication? The Interplay between Cognitive Ability and Social Origin in Explaining Educational Attainment, *European Sociological Review*. jcaa054.

van de Ham, A.-K., Schnittjer, I. and Gerken, A.-L. (2018). NEPS Technical Report for mathematics: scaling results of starting cohort 3 in grade 9, *NEPS Survey Paper No. 38*.