ISOTIS

INCLUSIVE EDUCATION AND SOCIAL SUPPORT TO TACKLE INEQUALITIES IN SOCIETY

Roots and Development of Achievement Gaps

A Longitudinal Assessment in Selected European Countries

EDITORS

Giampiero Passaretta and Jan Skopek

AUTHORS

Thomas van Huizen, Piergiorgio Lovaglio, Giampiero Passaretta, Luisa Antunes Ribeiro, Jan Skopek, Stefano Verzillo, Giorgio Vittadini, Henrik Daae Zachrisson

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT NO. 727069

Roots and Development of Achievement Gaps

A Longitudinal Assessment in Selected European Countries

EDITORS

Giampiero Passaretta and Jan Skopek

AUTHORS

Thomas van Huizen, Piergiorgio Lovaglio, Giampiero Passaretta, Luisa Antunes Ribeiro, Jan Skopek, Stefano Verzillo, Giorgio Vittadini, Henrik Daae Zachrisson

Document Identifier
D1.3

Version
1.0

Date Due
31st December 2018

Submission date
31st December 2018

Work Package
WP1

Lead Beneficiary
TCD

Citation guidelines

Citing the overall report:

Passaretta, G., & Skopek, J. (Eds.) (2018). *Roots and Development of Achievement Gaps. A Longitudinal Assessment in Selected European Countries.* ISOTIS Report (D 1.3), Trinity College Dublin.

Citing a specific chapter from the report (e.g., the chapter on the Netherlands):

van Huizen, T. (2018). The Evolution of Achievement Gaps from Early Childhood to Adolescence in the Netherlands. In: G. Passaretta and J. Skopek (eds.), *Roots and Development of Achievement Gaps. A Longitudinal Assessment in Selected European Countries, pp. 50-87,* ISOTIS Report (D 1.3), Trinity College Dublin.

Online Appendix

Several chapters in this report refer to additional and supplementary analyses which are provided in a separate appendix that is available online at http://www.isotis.org.

Change log

Version	Date	Amended by	Changes
1.0	18-12-2018	Jan Skopek, Giampiero Passaretta	Document finalized for submission
1.1	25-07-2019	Jan Skopek, Giampiero Passaretta	Merged reference list
1.2	14-05-2020	Jan Skopek	Minor revision introduction

Partners and researchers involved

Partner number	Partner name	Researchers involved
5	Trinity College Dublin (Ireland)	Jan Skopek (leader ISOTIS WP 1), Giampiero Passaretta
1	University of Utrecht (The Netherlands)	Thomas van Huizen
13	University of Oslo (Norway)	Henrik Daae Zachrisson, Luisa Ribeiro
7	University of Milano-Bicocca (Italy)	Piergiorgio Lovaglio, Stefano Verzillo, Giorgio Vittadini

Table of Contents

List of Abbreviations

ASQ:	Ages and Stages Questionnaire
BONDS:	The Behavior Outlook Norwegian Developmental Study
BPVS-II:	British Picture Vocabulary Scale-II
CCC2:	Children's Communication Checklist – 2nd rev
COOL:	Cohort Research on Educational Careers
ECEC:	Early Childhood Education and Care
EU:	European Union
INVALSI:	Italian National Institute for the Evaluation of the School System
IRT:	Item Response Theory
ISOTIS:	Inclusive education and social support to tackle inequalities in society
havo:	hoger algemeen voortgezet onderwijs [*senior general secondary educ.*]
K07/08:	2007/2008 kindergarten cohort
MCDI:	MacArthur Communicative Development Inventory
MCS:	Millennium Cohort Study
MoBa:	Norwegian Mother and Child Cohort Study
NEPS:	National Educational Panel Study
OECD:	Organisation for Economic Cooperation and Development
PIAAC:	Programme for the International Assessment of Adult Competencies
PIRLS:	Progress in International Reading Literacy Study
PISA:	Programme for International Student Assessment
Pre-COOL:	Cohort Research on Educational Careers – young child
RQ1:	Research Question 1
RQ2:	Research Question 2
SC1:	NEPS Starting Cohort 1
SC2:	NEPS Starting Cohort 2
SC3:	NEPS Starting Cohort 3
SD:	Standard deviation
SES:	Socio-economic Status
SSIS-RS:	Social Skills Improvement System Rating Scales
UK:	United Kingdom
WP1:	Work Package 1 of ISOTIS project
vmbo:	voorbereidend middelbaar beroepsonderwijs [pre-vocational secondary educ.]
vwo:	*voorbereidend wetenschappelijk onderwijs* [pre-university educ.]
z-score:	Standardized score variable (standard-deviation unit scale).

Executive Summary

This report presents results from ISOTIS Task 1.3 (Roots and development of skill gaps: A longitudinal assessment in selected European countries), a comprehensive longitudinal study of social and migration-/ethnicity related gaps in educational achievement in a European-wide comparative perspective. We analyse the evolution of achievement gaps in children from infancy and preschool age up to end of compulsory schooling in five country cases: Germany, Netherlands, Norway, the United Kingdom, and Italy. All country-case studies make use of recent high-quality cohort data to study in-depth two key research questions: (1) When do social and migration/ethnicity gaps in children’s achievement arise and how do they evolve when children are growing up and navigating from infancy to preschool, from preschool to school, and from primary to secondary school? And, (2), to which degree are social and migration-related inequalities in school-age achievement already determined by early inequalities well-established before children enter school? Taken together, we show that the early years of life (before children enter school) are formative for patterns of inequality observed in school age, and this holds for achievement inequality both by socio-economic and migration status. Socio-economic and migration-related achievement gaps in school are therefore rooted substantially in the early years.

Substantial inequality in educational achievements by the socio-economic status (SES) of children’s family of origin was found across all countries. Children from high-income families and from parents with a high level of education perform consistently better than children from less affluent families and whose parents have less educational resources. Importantly, these socially-determined gaps are already visible in the very early years of life, tend to increase steadily over infancy, and are well-established even before children enter primary school. After transition to school, SES-gaps in achievement remain quite stable and increase only slightly throughout years of primary and secondary education. Notwithstanding subtle differences across countries, we found considerable similarities in the evolution of socio-economically determined achievement gaps despite clear institutional differences in national education systems and overall welfare-state arrangements. Moreover, a major part of SES inequalities in achievement accumulated over the early years is carried over into the school system even though factors related to family SES continue to shape children's achievement in school. We conclude that preschool-age interventions that facilitate a more equalized start into school life hold the promise of reducing a large part of socio-economic achievement inequality in the later school career.

Our findings reveal more country heterogeneity regarding educational inequalities related to migration and ethnic minority background of children. In general, children with a migration or ethnic minority background enter school with a substantial disadvantage in achievement. Turkish and Moroccan children are particularly disadvantaged. Socio-economic disadvantages related to migration background could only in part account for those inequalities, yet these findings varied between countries and target groups. In some countries, however, initial disadvantages of migrant children vanish almost entirely after school entry. Extreme cases represent the UK (migrant gaps close quickly after school entry) and Germany (in general, migrant gaps do not decrease over schooling). For several countries we find that although children with a migration background are lagging behind at school entry, they enjoy over-proportional achievement gains in school life. Hence, when starting into school at the same achievement level, many migrant children are outperforming children of native families. Thus, reducing migration-related inequality in preschool-age could have the potential to eradicate migrants' penalties in school-age entirely.

1 INTRODUCTION

A Longitudinal and Comparative Study on Achievement Inequality in Europe

Giampiero Passaretta[♠] & Jan Skopek[♠]
♠Trinity College Dublin

1.1 Background

There is a broad consensus that family resources affect children educational achievements in virtually all European countries (Barone 2006; Marks, Cresswell, & Ainley 2006). However, while there is a wealth of studies on inequality in learning outcomes among children from different social and migration backgrounds, most of what we know restrict to specific age groups or educational stages, notably during secondary school age and adolescence. From a methodical standpoint, the existing research adopts cross-sectional research designs providing snapshots of inequality along the educational career, thus being unable to reconstruct the dynamics of educational inequality over the early life course.

By integrating cross-sectional data from various rounds of international assessment studies of students (PISA, TIMSS, PIRLS) and adults (PIAAC) into a pseudo-panel design in an earlier report (D1.2), ISOTIS Work Package 1 improved our understanding of the life course evolution of achievement (or skill) gaps in a variety of institutional contexts (Rözer & Werfhorst 2017). Nonetheless, similar to other research of this kind (see, for example, Dämmrich & Triventi 2018), the earlier report mostly focused on the dynamics of inequalities from end of primary school age (about age 10) towards adulthood and, due to the cross-sectional nature of the data at hand, could not provide a genuinely longitudinal account of the processes of inequality under study. Moreover, school assessment data is not suitable to analyse the early roots and evolution of children's diverging educational competencies.

Despite the focus on inequalities in educational achievement among adolescents, an accumulating body of research brought evidence that achievement inequality is rooted very early in children's lives. For example, previous research suggests that cognitive gaps in post-birth abilities among babies from varying social background are tiny in magnitude (Fryer & Levitt 2013) but subsequently proliferate when infants become toddlers and toddlers become preschool children (Feinstein 2003; Fernald, Marchman, & Weisleder 2013). At entry into Kindergarten, socio-economic gaps in early reading and math skills among US children are substantial (Bodovsky & Youn 2012; Lee & Burkam 2002). Hence, skill differentials by family background indeed may settle down very early in the educational career – virtually before school – thus challenging the role that education systems might play in the process of social stratification in Western countries. Therefore, social and ethnic inequalities in educational achievement observed by cross-sectional studies during adolescence are likely to reflect a complex cumulation of stratification processes that commence shaping experiences of children right from the earliest stages in childhood.

The exact point in which skills gaps emerge and how they evolve from infancy to adolescence is not well understood, however. On the one hand, we may expect skills gaps to appear very early and to increase over the educational career due to processes of cumulative disadvantage (DiPrete & Eirich 2006). On the other hand, we may also expect decreasing inequalities to the extent that education systems equalise conditions of learning to which children are exposed to (Alexander, Entwisle, & Olson 2004; Downey, von Hippel, & Broh 2004). Still, mechanisms of accumulation or compensation may work at the same time, and possibly resolve in the stability of initial inequalities over the educational career (Downey & Condron 2016; Skopek & Passaretta 2018).

A related issue is whether children with the same skill endowments at early life stages develop differently depending on their social and ethnic background. Family characteristics may play a role at later life stages of the educational career by compensating for lower performances at the start or boosting early achievements. Therefore, not only children from a disadvantaged background may suffer from a skill penalty since the very beginning of their education careers compared to children from better off families, but this penalty may also arise when there are no skill differences in the beginning. In other words, we can ask whether differences observed before the school entry fully explains achievement inequalities during the school years or whether the following disparities are also attributable to the role that the social and migration background may play over the school career.

Representative, comparative studies that trace when and how social and ethnic achievement gaps unfold over the early years are still scarce, particularly for Europe. Recent longitudinal research has analysed social gaps in achievement in the US, Australia, Canada and the UK (Bradbury et al. 2015; Caro et al. 2016; Feinstein 2003; Votruba-Drzal et al. 2015). Hitherto, no comparative study traces the evolution of social and ethnic achievement gaps for European countries which feature more heterogeneity in terms of the overall welfare state approaches, immigration regimes, structures of social inequality, as well as the organisation of early childhood education and care and schooling in the educational systems. This lack of knowledge is striking since it is the variation in educational gaps and trajectories across different states, systems, and regions that provide essential clues for identifying successful or poor strategies of educational policies and practices aiming to target and tackle social and ethnic inequalities in Europe efficiently.

1.2 Research objective and questions

The central aim of the following report from ISOTIS Working Package 1 (WP1), is to provide a comprehensive longitudinal study of social and migration gaps in educational achievements in a European-wide comparative perspective. The report is an integral part of the overall objective of WP 1 to study educational Inequality in various stages of the educational career (Skopek et al. 2017). More precisely, the report documents the work of ISOTIS Task 1.3 (Roots and development of skill gaps: A longitudinal assessment in selected European countries). While building upon the previous ISOTIS WP 1 report provided by Rözer & Werfhorst (2017), this report is complementing and extending the earlier one in several ways.

First, while Rözer & Werfhorst (2017) studied achievement gaps concerning the transition from primary to secondary schooling and from secondary schooling to adulthood, our report starts earlier by putting an explicit focus on gaps in children's cognitive development and achievement *before they enter school*. Our conceptual observation window to study the evolution of achievement gaps entails four institutional stages of the early educational life course: *infancy and* *toddlerhood*, *preschool age*, *primary school age* up to *secondary school age*.

Second, in contrast to the earlier report who had built up a rich database of international but cross-sectional data, we are drawing upon truly longitudinal data involving repeated measurement on the *same* children. It is the longitudinal nature of our data that facilitates a more dynamic study of the evolution of achievement inequality. Repeated measures over preschool and school age not only allow to provide a sharper picture of the temporal development of achievement gaps but also enable to take into account the individual history of unequal achievement at each stage of the early life course.

Third, while the earlier report accomplished an extensive overview of achievement gaps by exploiting cross-national variation from a multitude of countries within Europe and across the world (in total 103 regions), this report goes more in-depth into the specific case of five selected European countries. All five countries are included in the overall comparative design of the ISOTIS main interview study (Broekhuizen et al. 2018) – Germany, Italy, Netherlands, Norway, and the United Kingdom – and their selection entered not only pragmatic reasons (such as the availability of high-quality longitudinal data and overlap with the overall ISOTIS design) but also guided by theoretical considerations. The chosen countries represent a fascinating set of diverse societies in terms of general welfare state approaches and underpinning political ideologies (e.g., liberal versus social-democratic), concrete institutional organisation of preschool and school education (universal versus non-universal preschool, comprehensive versus tracked schooling, centralisation versus subsidiarity principle), the general socio-cultural and economic fabric, and different immigration histories and policies (skill-based versus low-skilled/humanitarian immigration). Furthermore, the in-depth strategy based on high-quality longitudinal taken adopted by this report allows a finer distinction of target groups pertinent to the overall ISOTIS approach (e.g., low-income families, Turkish, or other ethnic minority groups) which was infeasible for the earlier report working with international data.

Hence, our approach can add more details and nuances to the study of educational inequality particularly in the early stages of the educational career in the specific societal, institutional, and educational context of these countries. Moreover, this approach permitted us to exploit the best and recently available longitudinal data sets (mostly collected by representative and large-scale cohort studies) that were available for the selected country cases. Such level of detail, however, comes at some expenses in terms of comparability as – unlike PISA data – country-specific datasets have not been collected in a centralised way and feature partly different cohorts of children, slightly different sampling approaches, and partially different measures and follow up windows. To minimise issues of comparability, all country-studies adopted a common research design that has been developed before analyses were carried out and specified main research objectives, research questions, methodology, and coding procedures that were equivalent as far as possible. We defined two broad sets of research questions which will be specifically addressed in the country chapters:

RQ1 – When do social and migration/ethnicity gaps in children’s achievement arise and how do they evolve when children are growing up and navigating from infancy to preschool, from preschool to school, and from primary to secondary school?

RQ2 – How predictive are preschool inequalities for later inequalities in school? How much of social and migration/ethnicity gaps in educational achievement observed later in school is rooted in achievement disparities before children had entered school? Does social and migration/ethnicity background play a role in shaping inequalities in school life beyond the preschool period?

Answers to the first set of research questions will provide better knowledge about the size and overall time evolution of achievement gaps. For effective educational policies, it is of particular relevance to gather better time-related information about the strength of social and ethnic inequalities at various stages of early childhood and school life as well as the life stages in which achievement gaps are emerging, widening or potentially even reducing. Our report is the first attempt in the European literature that integrates longitudinal evidence on how educational inequalities develop from infancy to secondary schooling age along the lines of family socio-economic status and migration/ethnicity for several European countries. More specifically, we answer the following questions: *How large are social and migration gaps in educational achievements at different stages of the educational cycle? What is the size of the skill gap before school entry? How do gaps develop during the school years and particularly during the transitions from pre-primary to primary and secondary schooling? Is the evolution of inequality similar or different across institutional contexts?*

The second set of research questions calls for truly longitudinal analyses in which achievement gaps at later educational stages are conditioned on early inequalities before school. Answers to the second question will give us an empirically more accurate picture of the relative strength of inequalities processes that operate before children transit to school and those that operate after children's transition to the school system. More precise knowledge on the relative importance of preschool and school processes can inform educational policies about the timing of interventions, that is *when* in children's lives it is most crucial to combat inequalities and *which children* should be the targeted primarily. More specific questions in the second set of questions are: *To what extent do social and migration gaps in early skills translate into disadvantages in school? Is the role of preschool inequality in shaping later disparities similar or different across the institutional contexts? Does family background play a role in shaping school inequalities beyond the preschool period? Is the additional role of family background over schooling concentrated among children performing high or low in preschool?*

1.3 The relative approach

A study of the evolution of achievement gaps over an extended lifespan entails substantial challenges in measurement. One of the most obvious issues is that children master different sets of skills at different times due to the hierarchical nature of the process of development itself. For example, the attainment of math skills in school requires the ability to read and learning to read entails the ability to comprehend words and sentences, an ability to be mastered before school. What is more, developmental changes in the early years of life are exceptionally rapid and complex (Zeanah, Boris, & Larrieu 1997). The qualitative transformations inherent to the developmental process itself renders the measurement of the early life course evolution of skills infeasible on absolute scales (Feinstein 2003). A related pragmatic challenge roots in the fact that most longitudinal cohort studies providing the database for this report did test a variety of skill domains and, although several core domains were repeatedly measured, this had not occurred in each successive wave.

In line with previous longitudinal studies (e.g., Bradbury et al. 2015; Feinstein 2003;

Hoffman 2018) and the previous WP 1 report (Rözer & Werfhorst 2017), our measurement approach relied on the construction of *relative test score differences* rather than absolute differences on the tests' scale. Relative measures are favourable for studying inequality as they express test score gaps in relation to the overall test score variation and therefore measure inequality in the distribution of achievement across groups (Reardon 2008). In practice, test scores are *z-standardised* within each time point and express the relative differences between children in the same life stage. The standardisation imposes an overall mean of 0 and a standard deviation of 1 in the *z-score* – our relative measure –, thus increasing the comparability of achievement gaps across life stages. Hence, the use of relative measures is a viable strategy to study the evolution of inequalities in achievement over an extended time window. Moreover, the relative approach is particularly suited to the purpose of our study, as we are interested in the stratification of skills among groups and not in the process of development itself. Last but not least, using relative measures establishes compatibility with the earlier multi-national assessment of ISOTIS WP 1 which adopted the same approach to analyse educational inequality from school to adulthood.

A relative measurement approach brings many advantages in the context of our temporally and cross-national comparative study but also imposes several limitations. Most importantly, relative skill differences – in terms of *z-score* – and absolute skill levels – in terms of a test specific proficiency scale – should not be confused. Relative measures express inequality while absolute proficiency or competence measures may express the specific process of solving a test (e.g., in form of a sum score) or some associated latent ability. Hence, the evolution of standardised scores along the life course does not say anything about children's competence development in any absolute sense and, as a consequence, z-scores cannot be used to analyse individual growth in a skill domain (which is not our aim here). Nor can z-scores be used to gauge the actual competence difference between groups in a specific domain. Moreover, by their nature of measuring inequality, z-scores are a function of the overall variability of the underlying test scores or competence measures. As a consequence, it is possible that mean differences in the z-scores decrease over time as a result of increasing heterogeneity (variance) in absolute competencies, although absolute mean differences in competencies remain constant.[1]

1.4 Dimensions of stratification

Family socio-economic status (SES) and migration experience are two essential drivers of achievement inequality in children and students of Western societies. SES refers to the social position of the family of origin in a stratification system, which includes access to financial and material resources, skills and knowledge, and social capital (Bradley & Corwyn 2002; Duncan & Magnuson 2003; Oakes & Rossi 2003). Two of the most critical dimensions of family SES are

[1] There are two additional limitations. First, like absolute scores, standardised scores hinge on the assumption of perfectly interval scaled test score variables. This assumption is relaxed by metric-free measures that have purely ordinal features such as percentile ranks or relative distribution measures (Reardon 2008) that, nonetheless, make less efficient use of the data if it is actually interval scaled. In our case, sensitivity analyses using percentile ranks yielded similar results. Second, the average standardise score across groups are sensitive to the marginal distribution of those groups, although of no consequence for between-group differences in the average standardised score (the real quantity of interest in our context). Intuitively, this occurs because the relative groups' size influences the average standardised score within groups, an issue not explicitly discussed in the existing literature. Note, however, that the problem is only pertinent when using unbalanced panels of children (the marginal distributions are constant by construction in balanced panels). Each of the country chapters adopts specific strategies to tackle this issue.

parental education and household income and, while interrelated, each of them governs distinct mechanisms of inequalities (Duncan & Magnuson 2003; Duncan, Magnuson, & Votruba-Drzal 2015). On the one hand, high parental education translates into better material, cultural and social resources that are available to support child development (Conger & Donnellan 2007). Better educated parents are likely to have more knowledge and skills that can be transmitted to children directly, through everyday interaction, or indirectly, through parenting practices (Ermisch 2008) or stronger involvement in children's life (Domina 2005; Fan & Chen 2001). On the other hand, high-income families may provide children with outstanding material resources at home, high-quality child care, access to good schools, or private tutoring. Moreover, experiences of economic hardship entail emotional distress for parents, with negative consequences on their parenting and on family functioning in general, which may have detrimental effects on children's cognitive development (Conger & Donnellan 2007; McLoyd 1998).

There are also significant differences in the conceptualisation of parental education and household income as stratifying dimensions when we adopt a life course perspective. On the one hand, parental education is a quite stable measure of social background since the proportion of individuals upgrading their education level after childbirth is relatively low in many European countries. Conversely, household income is a less stable and more volatile measure of background since occupational progression seems to persist until 10–15 years after school-leaving in many institutional contexts, such as the UK (Bukodi & Goldthorpe 2011), Germany (Manzoni, Härkönen, & Mayer 2014), Italy and the Netherlands (Passaretta et al. 2018). Therefore, parental education is a more stable indicator of SES in our context. Nonetheless, this report details the analyses for both parental education and household income to offer a complementary view of the mechanisms governing the role of SES in the stratification of educational achievement. We adopt a categorical view on inequalities by measuring both parental education and household income as groups rather than continuous variables, in line with the overarching ISOTIS strategy of focusing on marginal groups in the society. However, depending on theoretical and practical considerations, some country studies may slightly deviate from this common strategy.

Immigration experiences of a family is another critical dimension that may lead to educational disadvantage. Children with a migration background often find themselves in lower SES families compared to the average child in the host country. Hence, differences by SES may itself explain in part potential differences in the educational performance of children of immigrants and children of natives. The interaction between origin and destination countries renders educational disadvantage related to migration background a contextual phenomenon that varies across countries (Levels, Dronkers, & Kraaykamp 2008). For example, children with a migration background may have more similar or even higher SES background than non-migrants in countries with selective migration policies (like the UK) compared to countries with a long-standing 'guest worker' (like for example Germany, Netherlands or Italy) or 'humanitarian' tradition (like Norway). Next to such compositional differences between immigrant communities in various countries, other factors related to the specific countries of origin may matter.

Apart from socio-economic differences between migrant and non-migrant families, the migration background itself may play a significant role in children's educational adjustment. For example, children of immigrants may experience educational disadvantages insofar as their parents do not master the host language and use the native language for everyday interactions at home (Crozier and Davies 2007; van de Werfhorst & van Tubergen 2007). In this scenario, children with a migration background would experience an impairment of their host language skills and, to the extent that the host language is critical for interactions with other children and teachers at school, also different sets of skills would be equally impaired (the mastery of language is essential for reading a textbook or asking clarifications to teachers, for instance). Moreover, children of immigrants are not socialised to the education system of the host country, and this may lower their motivation to learn or even exposed them to discriminatory behaviours from the side of teachers and classmates (van de Werfhorst & Hofstede 2007). Worth noting is that even in the case of disadvantages directly connected to migration background, significant differences may exist depending on the host and sending countries. For example, stronger cultural proximity between origin and host countries is likely to results in lower differences between the performances of migrants and non-migrants since children with a migration background will to some extent be accustomed to the core values and the structure of the hosting education system. In this report, we examine differences between the educational achievement of children with a migration background (at least one parent born abroad) and non-migrants in a variety of host countries.

Next to immigrant background, we also inspect difference by ethnicity, which in some countries is a more prominent issue than migration background. As far as the data permitted, we aimed to address some of the ISOTIS target groups specifically (e.g., Turkish immigrants). Unfortunately, data limitations did not allow us to detail the analyses for other ISOTIS target groups (such as Roma). Nonetheless, based on country-specific considerations, we extend the analyses to other important migrant groups to locate the ISOTIS target groups in a country's larger context of immigration.

1.5 The five countries under study

Our longitudinal analysis of social and migration gaps in educational achievements is embedded in a comparative framework. We selected five European countries that show a remarkable variety in their welfare arrangements, levels of inequality and, above all, the organisation of education systems: Germany, the Netherlands, Norway, the United Kingdom, and Italy. Table 1 provides an overview to the country cases. The selection of these five countries ties with the ambition of providing a comprehensive portrait of European societies, although partly based on the availability of suitable longitudinal data at the national level.

We selected the United Kingdom as a benchmark, as it resembles the institutional configuration of the classical 'liberal' Anglo-Saxon countries analysed within a life course perspective so far (notably the United States, Canada, and Australia). Most recent and high-quality longitudinal data from the Millennium Cohort Study are used. While already been targeted by longitudinal research on achievement gaps (e.g., Bradbury et al. 2015; Hoffman 2018), we provide new insights on the UK case study by exploiting in full the longitudinal component of the data, focusing on a broader time window, analysing several dimensions of SES, migration and ethnicity, and adopting a revised strategy for tackling panel attrition to ensure the representativeness of the analyses at the national level (see the UK country chapter for details).

Table 1 Overview to the country cases – institutional features, data, and observation windows.

	INSTITUTIONAL FEATURES					DATA	
	Welfare regime	**Income inequality**	**ECEC**	**Primary education**	**Secondary education**	**Survey**	**Obs. window (age)**
GERMANY	Conservative	Medium-Low	*Age 0–3* • Public-private mix • Low availability • Low public spending *Age 3–5 'Kindergarten'* • Nearly universal	Comprehensive (age 6)	Early tracking (age 10, partly ability-based)	National Educational Panel Study (NEPS)	0 – 15/16
NETHER-LANDS	Conservative/ Social-democratic	Medium-Low	*Age 0–4* *'Split' ECEC system* • Day-care (0–4) • Preschool (2.5–4), *Age 4–5* • Universal	Comprehensive (age 6, includes Kindergarten 4–5)	Early tracking (age 12, ability-based)	Cohort Research on Educational Careers in The Netherlands (PreCOOL and COOL)	2 – 14
UNITED KINGDOM	Liberal	High	*Age 0–3* • Private market • Low availability • Low affordability *Age 3–4* • Free part-time education	Comprehensive (age 5)	Comprehensive (no formal tracking) Curriculum/school differentiation	Millennium Cohort Study (MCS)	3 – 14
NORWAY	Social-democratic	Low	*Age 1–5* • Universal access • High public spending • High quality	Comprehensive (age 6)	Late tracking in upper secondary schooling (age 16, partly ability-based)	The Behavior Outlook Norwegian Developmental Study (BONDS) Norwegian Mother and Child Cohort Study (MoBa)	0 – 8
ITALY	Conservative/ Mediterranean	High	*Age 0–3* • Public-private mix • Low availability • Low public spending *Age 3–6* • Nearly universal	Comprehensive (age 6)	Late tracking (age 15, not ability-based)	Italian National Institute for the Evaluation of the School System (INVALSI)	10 – 15

In addition to the UK, we consider four additional countries from the full spectrum of European welfare regimes: Norway as a representative of the social-democratic model, Germany as a conservative welfare state, the Netherlands as a hybrid between the conservative model and the flexicurity arrangement typical of Scandinavian countries, and Italy as an example of the Mediterranean branch of the conservative cluster (Esping-Andersen 1990; Muffels & Luijkx 2008). In this way, we contrast the European country with the highest level of income inequality and a residual role of the state in the economic and societal spheres – that is the UK – with the prototypical European contrast showing the lowest level of income inequality and a significant role of the state in the decommodification of citizens from market dynamics – that is Norway. In between these two extremes, we contrast three different shades of the conservative arrangements resulting in moderate (the Netherlands and Germany) and relatively high (Italy) levels of inequality among the family environments (Smeeding & Rainwater 2004). Moreover, the countries under study feature different traditions in immigration flows (OECD 2018a): UK has a longstanding immigration with highly educated migrants; Germany and Netherlands share a long-standing (guest worker) immigration tradition with typically lower educated labour migrants; Italy represent a more recent migrant destination characterised by low-education labour migrants; and, finally, Norway experienced a more recent mainly humanitarian immigration.

Apart from the logic behind the overall welfare arrangements and immigration policies, the five countries differ considerably also in the organisation of their education systems. In all countries, preschool is organised differently for children in the early years (0–3, approximately) and the period immediately preceding the commencement of primary education (4–6, roughly). Despite the significant public spending in the UK, parents' can only rely on costly and scarcely available private ECEC arrangements in the early years, while being offered just free part-time education for children aged 3–4. The UK system contrasts sharply with both the Dutch system – combining various ECEC arrangements with relatively high attendance rates (but low intensity) and universally available Kindergarten (Leseman et al. 2017) – and the universal, high-quality and highly subsidised system of preschool education in Norway (Zachrisson et al. 2017). Finally, Germany and Italy stand in between by combining semi-standardised and scarcely subsidised early childhood education and care (ECEC) and standardised Kindergarten with nearly universal attendance (Blossfeld et al. 2017). Worth noting is a peculiarity of the Dutch system of preschool education. The Dutch Kindergarten is formally part of primary education, although being similar in its pedagogic approach to other systems that embeds Kindergarten in the welfare arrangement, such as (Oberheumer and Ulich 1997). Therefore, in the Netherlands, schooling formally starts earlier compared to all the other countries under study (age 4 vs age 5/7, approximately). Despite this subtle difference in the starting ages, primary education is standardised and comprehensive in all five countries, although there is some degree of informal differentiation through social and ethnic/migration-related segregation of schools.

The five countries under study vary significantly in the organisation of secondary education particularly in formal and informal ways of educational differentiation (Blossfeld et al. 2016). Germany and the Netherlands are both examples of early tracking models which track students to different school types (typically academic track, medium track, lower vocational track) at age 10/11 and 12/13, respectively. However, while track allocation is only partially based on demonstrated previous abilities in Germany (in some German states based on a binding primary school teacher recommendation), ability-based sorting is more pronounced in the Netherlands (secondary school track placement is based on primary school recommendations and elementary exit tests). In contrast, Italy and Norway are good examples of late tracking (age 14/15 and 16, respectively). However, the two systems differ in the criteria behind the tracking of students. In Italy, previous performances do not constitute a formal criterion for tracking students in different educational pathways, while in Norway tracking is partly based on earlier grades although, in practice, almost all students enter their preferred track. The only exception to tracking is the UK, which is mostly characterised by a comprehensive approach in secondary schooling. Nonetheless, the lack of formal tracking couples with extensive heterogeneity across secondary schools, curriculum differentiation by subject choice, and overall school quality.

Taken together, the five selected countries exhibit a significant variation not only in their overall welfare state arrangements but also, more specifically, in their institutional organisation of early childhood and care and schooling. Therefore, the comparative approach of our study is well suited to portrait the evolution of educational inequality in the contextual heterogeneity of European countries. A cross-country comparison of inequality patterns will shed light on the role of critical institutional characteristics possibly impacting on the formation and the evolution of social and migration inequalities in educational achievement over the life course of children.

1.6 Outline

The remaining of this document is organised in five country chapters and a general concluding section. While embedded in a common theoretical and analytical framework, each chapter is authored by each of the country-teams and can be read independently. The country chapters start by putting the general research questions (formulated above) into context and describing the relevant institutional characteristics with a particular emphasis on the role of national education systems. Then, each chapter describes the strengths and weaknesses of the data at hand and provides details about the methods and all statistical procedures involved. After reporting the results, the chapters conclude by outlining and interpreting the main findings in light of the institutional setting.

While answering the same set of research questions (RQ1 and RQ2), each chapter focuses on slightly different aspects depending on country-specific considerations as well as strengths and limitations of the data at hand. Using most recent and high-quality cohort data from the *German National Educational Panel Study* (NEPS), the Dutch *Cohort Study of Educational Careers* (COOL and Pre-COOL), and the UK *Millennium Cohort Study* (MCS) the chapters on Germany (contributed by Passaretta & Skopek), Netherlands (contributed by van Huizen) and the UK (contributed by Skopek & Passaretta) focus on an unprecedentedly large observation windows ranging from the early years until the end of compulsory education (see Table 1). Moreover, in these three countries, the authors can focus on multiple competence domains and sometimes even a composite measure of achievement, thus exceeding the scope of previous research in significant ways.

There are some limitations too. For example, in the case of Germany, the authors could not analyse some ISOTIS target groups – such as the Maghrebs – due to lack of detailed information on children's migration background. What is more, the small sample size in both Germany and the Netherlands did not permit detailed analyses by specific groups of migrants when focusing on the role of preschool inequalities for achievement disparities in school (RQ2). The MCS data provided a large sample of children that allowed very detailed sub-group analyses for the UK. Although information on ISOTIS target groups was not available, the UK chapter highlights educational inequality in the context of ethnic minority groups which are important target groups of educational policies in the UK context (such as Black African/Caribbean, Pakistani and Bangladeshi, or Indian).

Data limitations dictated shorter observation windows in the case of Italy (age 10–15) and Norway (age 6 months to 8 years). The Italian chapter (contributed by Lovaglio, Verzillo & Vittadini) employed full population data from the *Italian National Institute for the Evaluation of the School System* (INVALSI). Population data on more than 500,000 students limited the uncertainty around the empirical results substantially and allowed the authors to investigate subtle differentiations by migrants' generation status (first and second) and gender. As a drawback, the Italian chapter had to use a pseudo-panel approach because the actual longitudinal component of the INVALSI data was not available to date. Furthermore, there was no detailed information on migrants' country of origin available in the Italian data. For Norway (contributed by Zachrisson & Ribeiro), the lack of data after age eight is counterbalanced by high precision in the early years of life. Two mutually complementing datasets from the *The Behavior Outlook Norwegian Developmental Study* (BONDS) and the large-scale *Norwegian Mother and Child Cohort Study* (MoBa) have been employed. BONDS is a smaller study with a wide array of measures but focused on specific Norwegian regions. In contrast, MoBa is a nation-wide study with approximately 100,000 children included.

The document concludes with a final section in which we review the main findings from the country chapters with the aim of identifying broader lessons learned and recommendations for policy.

2 GERMANY

From Birth to the End of Compulsory School – Social and Migration-related Achievement Inequality in a Stratified Education System

Giampiero Passaretta♠ & Jan Skopek♠
♠Trinity College Dublin

2.1 Introduction

The results of the first PISA assessment in 2000 came as a shock in a country that had viewed itself as an educational model-to-follow for many years (Ertl 2006; Gruber 2006; Ringarp & Rothland 2010). German's students performed below the OECD average in many competence domains including reading, math and science. Moreover, differences in the performances of students astoundingly aligned along the lines of social and migration background (Deutsches Pisa-Konsortium 2001). The following round of PISA in 2003 only saw a slight improvement in the average performance that, nonetheless, was mostly driven by those students enrolled in the higher track of secondary education (Klein 2005), where children from lower social strata and migration background are traditionally underrepresented. Therefore, if anything, the slight improvement in the overall performance came at the expense of higher inequalities in educational achievements.

The PISA shock activated a public discourse in the German society around the efficiency and the equality of the education system, which culminated in the introduction of national educational standards and the support of disadvantaged students (Ertl 2006). In this sense, studies such as PISA has proven key in orienting the political, societal, and academic discourse about institutional change in the educational sphere. However, as argued in the introductory chapter, these cross-sectional studies only offer snapshots of inequalities during adolescence, thus concealing the dynamics of inequalities as they unfold over infancy and childhood. The lack of longitudinal accounts is surprising in light of the German education system, which is argued to be one of the most stratifying in the Western World. In Germany, educational and occupational destinies begin to take shape early in life as children are tracked in very different educational pathways at young ages. Therefore, it is particularly interesting to examine how inequalities in educational achievement unfold and evolve over the life course of children in the German context of education. We try to answer both the research questions outlined in the initial chapter of this report.

RQ1 – First of all, we ask how social and migration inequalities in the educational achievements unfold and evolve over the early years of life until the end of lower secondary education. More precisely, we aim at answering the following questions:

(1) When do social and migration gaps in cognitive achievement emerge?
(2) How large are the gaps before children enter school?
(3) How do these gaps develop during primary and secondary schooling?

RQ2 – Second, we ask whether and the extent to which social and migration inequalities in later educational stages (primary schooling) are explained by inequalities already settled before the entry into the education system (the preschool period). More precisely, we aim at answering the following questions:

(1) Are social and migration gaps in primary school fully explained by preschool inequalities?
(2) What is the role of social and migration background in shaping inequalities beyond the early years?
(3) Is the role of social and migration background over schooling concentrated among high or low performing children in kindergarten?

The socio-economic status (SES) and the migration background of children are two of the most critical dimensions shaping educational inequality in Germany, as suggested by the public discourse in the aftermath of the PISA shock. Regarding SES, we can detail the analyses by both parental education and household income.

Regarding migration background, we focus on the overall differences between the educational performances of native and migrants, on the one hand, and on the differences between the performances of natives and specific groups of migrants, on the other. Unfortunately, the German data allows us to focus only on one particular group of migrants targeted by ISOTIS, that is Turkish. Nonetheless, we complement the focus on Turkish by analysing a different but also prominent group of migrants in Germany, that is Russians. Turks and Russians have a long-lasting, well-establish tradition of migration to Germany and are currently the most prominent groups of migrants in the country. While sharing a common background of immigration, children from the two groups differ considerably in their cultural heritage, with children from Russian background being more similar to Germans compared to children with a Turkish origin. Hence, the comparison between non-migrants and these two specific groups offers valuable theoretical insights into the causes underlying migration inequalities in educational achievements.

The analyses rely on data from the German National Education Panel Study (Blossfeld, Roßbach, & von Maurice 2011). The NEPS is the largest national assessment dataset in Europe in terms of richness of the longitudinal design and variety of competence domains tested. These strengths allow us to focus on the evolution of achievement gaps over an exceptionally extended time window – from birth to age 16 – and in a wide array of competence domains. Moreover, a multitude of competence assessments allows us to construct a single composite measure of achievement that offers a straightforward and synthetic portrait of the evolution of inequalities over childhood.

NEPS data have some significant limitations, however. First, the limited information about the migration background of children does not allow us distinguishing other target groups in ISOTIS, such as Magrebian and Roma. Second, the longitudinal component of the data is restricted to cohorts of children followed starting from different points in their life course. Moreover, there is a significant number of children who drop out the study over time. These latter issues are addressed by adopting a careful weighing strategy that allows us to link different cohorts of children and account for selective attrition over time (see Section 2.3.1). Still, genuine longitudinal analyses (RQ2) are only possible when looking on a limited time window (within cohorts).

The most critical limitation is, however, the restricted sample size. The problem is particularly relevant when analysing RQ2, which requires truly longitudinal data. For example, the low number of migrants in the sample does not allow us to detail the analyses by the specific migrants' groups (Russians and Turkish). Moreover, when analysing the contribution of preschool inequalities for school inequalities, we are forced to focus on parsimonious metric measures of SES rather than low-educated and low-income families. Finally, the low sample size significantly increases the uncertainty around our results. As a consequence, some of the findings related to RQ2 have to be conceived as simulations rather than strict descriptions of the processes of inequality under study.

The remaining of the chapter is organised as follows. In the next section, we describe the main characteristics of the education and the family context in Germany and briefly discuss their implications for social and migration inequalities in achievement. The third section presents the data, the variables and the methods used to examine both research questions. In the fourth section, we show first results describing the evolution of achievement inequality (RQ1) and then results explaining inequalities in primary school by preschool inequalities (RQ2). In both subsections, we examine first inequalities by SES and then inequalities by migration background. We conclude the chapter with a summary of the findings and a discussion of the main results connecting to the education and the family contexts in Germany.

2.2 The German context

2.2.1 The family environment

Germany embodies the model of 'conservative' welfare state characterised by a system of comparably generous welfare benefits, preservation of status hierarchies, and a marked principle of subsidiarity in social policy (Arts & Gelissen 2002). The German context differs starkly compared to the 'liberal', the 'Southern', and the 'socio-democratic' welfare state models typical of Anglophone, Mediterranean, and Scandinavian countries. At the same time, the German model is somewhat similar to the conservative welfare regime of the Netherlands.

The contextual differences in social policy and welfare arrangements may be consequential for the production of achievement inequalities. Income inequality and social disparities in living conditions in Germany are similar to the Netherlands but higher compared to Scandinavian countries (Norway) and lower compared to both liberal welfare regimes (the UK) and Southern European countries (Italy) (Smeeding & Rainwater 2004). Inequality in income and the living conditions among families are relevant as they may translate into parenting quality through mechanisms of family investment and stress. Insofar as heterogeneities in the living conditions map along ascriptive characteristics such as social and migration background, we can expect social and migration disparities in cognitive development in Germany to be higher compared to Norway, lower compared to the UK and Italy, and similar compared to the Netherlands.

2.2.2 The education system

2.2.2.1 Early childhood education and care

Early childhood education and care (ECEC) in Germany is framed within the child and youth welfare system and is not part of formal education. The disconnection with the school system is apparent in the title of the professional staff involved: educators rather than teachers. However, notwithstanding the formal separation with school, ECEC in Germany is based on a social pedagogy approach in which care and education are seen as inseparable components (Oberheumer and Ulich 1997). The system is organised around the principle of subsidiarity, and civic corporations have priority over local government in the social organisation of the service. Worth noting is that Catholic and Protestant churches are among the most significant providers: together with other welfare organisations, churches provide around the 60% of the places available (Deutsches Jugendinstitut Dortmunder, Arbeitsstelle Kinder- und Jugendhilfestatistik 2008).

While the ECEC is designed by law to assist parents in reconciling work and family, care services are often offered on a half-day basis which forces parents to look for additional help through private arrangements (Spiess, Büchel, & Wagner 2003). The ECEC system is substantially different for children in the early years – until age 3 – and children between age 3 and the formal start of schooling. In the early years (birth to age 3) care services are semi-standardised and, although rates of enrolment are rising, there is no universal provision (BMFSFJ 2015). Moreover, recent research underlined the social selectivity in the access to ECEC: children from higher SES families enjoy more exposure to early care and education institutions compared to children from low SES families, although there is significant regional heterogeneity (Skopek 2017).

Kindergarten is still the most significant preschool institution for children aged between three and six in Germany. While families may opt for family day care services, this is a considerable alternative only among children below age three (Leu & Schelle 2009). While not being mandatory and officially providing both care and education, kindergarten is intended to prepare children for school and have a nearly universal attendance rate of more than 90 per cent (OECD 2015). However, notwithstanding the coverage, German kindergartens differ considerably in the social and ethnic composition of children (Becker & Schober 2017).

2.2.2.2 Primary education

In Germany, children must enter primary education (*Grundschule*) the year they turn six by law. Participation in primary schooling is compulsory and, in most states, it covers grades one to four, that is until age 10 approximately. Despite some regional heterogeneity from state to state, primary education is highly standardised in terms of organisational structures, curricula, and teachers' professional training (Allmendinger 1989). Therefore, children are exposed to the same type of instruction throughout the four years of primary schooling. It is likely that the highly standardised environments in German primary schools exert some compensation on achievement inequality. For instance, Baumert and colleagues (2012) studied social and migration inequality in learning outcomes and found compensation effects in reading skills to the benefit of primary school students with a migration background (increasing inequality, though, was found for math).

2.2.2.3 Secondary education

The German system of secondary education is famous for being a strongly stratifying sorting machine (Allmendinger 1989; Blossfeld et al. 2016; Bol & van de Werfhorst 2011). Nearly all of the 16 German states track students in the first year of secondary education (Grade 5), when they are aged approximately 10–11 years old. The three major school tracks – academic (*Gymnasium*), poly-technical (*Realschule*), and lower vocational (*Hauptschule*) – embody different curricula and by and large different school types. The academic is the most demanding track and traditionally offers access to higher education and prestigious occupations. The poly-technical track, instead, is less demanding and provides access to clerical occupations. In the end, the lower vocational track has the lowest requirements and often leads to labour market opportunities limited to manual jobs.

Despite complex regional heterogeneity in the organisation and structure of tracking (Helbig & Nikolai 2015), the tripartite system is deeply rooted in the socio-historical context of Germany (Benavot & Resnik 2004). A common element across German states is the primary school recommendation (Gresch, Baumert, & Maaz 2010; Neugebauer 2010): at the end of primary school, teachers provide a formal recommendation for students stating the secondary school track most appropriate to their abilities (usually marks in math and German), behaviour, and talent. Specifically, the recommendation indicates children's eligibility for the academic track. The regulating power of the recommendation and its consequences, however, varies by federal state, from being just a suggestion to the parents to have legally binding character (Helbig & Nikolai 2015).

The previous research found considerable social and migration-related inequalities in the allocation to different tracks at the transition to secondary education (for example, Buchholz et al. 2016; Kristen & Dollmann 2010). The inequalities in track choice are to a significant extent – but not entirely – explained by students' educational achievement before tracking. Nonetheless, students from more advantaged social backgrounds are more likely to enrol to the academic track even when compared to disadvantaged students with the same pre-tracking achievement. What is more, after the initial allocation, students from high SES background are also more likely to experience upward mobility in track-type while low SES students are more likely to experience downward mobility (Buchholz et al. 2016; Schneider 2008; Stocké 2007). The previous findings suggest that differences in previous performances and the SES background entirely explain the lower enrolment rates of migrants' children to the academic track compare to natives. Indeed, migrant kids are even more likely to attend the academic track when compared with natives with similar previous performances and family SES (Kristen & Dollmann 2010).

2.3 Data

2.3.1 The National Educational Panel Study

NEPS data are particularly suited to our purpose as they include both information on the competence development of children and the family environment. The study has a multi-cohort sequence design in which several nationally representative cohort samples of children – 6 in total – are followed up and tested yearly on a series of competence domains starting from different points over their lives. In this report, we used data from three starting cohorts: the *Newborns Cohort (SC1),* the *Kindergarten Cohort (SC2)*, and the *Grade 5 Cohort* (SC3). Detailed information on the three starting cohort used (target population, sample size, and approximate age of first competence assessment) can be found in Table 1.

Table 1 Details on the three cohort samples of children. NEPS data.

	Target population	Sample size	~Age 1st assessment
SC1 – Newborns	Children born first half of 2012	3,481	6 months
SC2 – Kindergarten	Children attending kindergarten in 2010–11	2,996	5 years
SC3 – Grade 5	Children attending grade 5 in the school year 2010–11	6,112	11 years

Notes: SC2 and SC3 were supplemented by refreshment samples in Wave 3 (N=6,341 for SC1 and N=2,205 for SC3).

Data from the three cohorts link through a weighting approach that ensures the representativeness of the data at the national level and accounts for possible bias due to study dropouts and disproportion in the marginal distribution of the key variables across starting cohorts. A detailed description of the weighting procedures is given in the Appendix (see Appendix 2.1).

This strategy allowed us to focus on disparities in the cognitive development of children over an unprecedentedly extended time-window ranging from infancy – 7 months of age – to the end of compulsory schooling – age 15–16. Figure 1 shows graphically the structure of our linked longitudinal design, which includes in total more than 50,000 yearly-observations from an unbalanced sample of more than 15,000 children observed over 15 waves.[1]

The description of the evolution of achievement gaps (RQ1) relies on the full unbalanced sample. The explanation of school achievements gaps by preschool achievements (RQ2) strictly requires observing the same children moving from preschool to primary education and, therefore, relies on the balanced panel of children in SC2 only. For both research questions, we use slightly different samples in each step of the analysis to maximise statistical power.

[1] We excluded Wave 2 of SC1 and Wave 4 of SC3 because no appropriate test data was available.

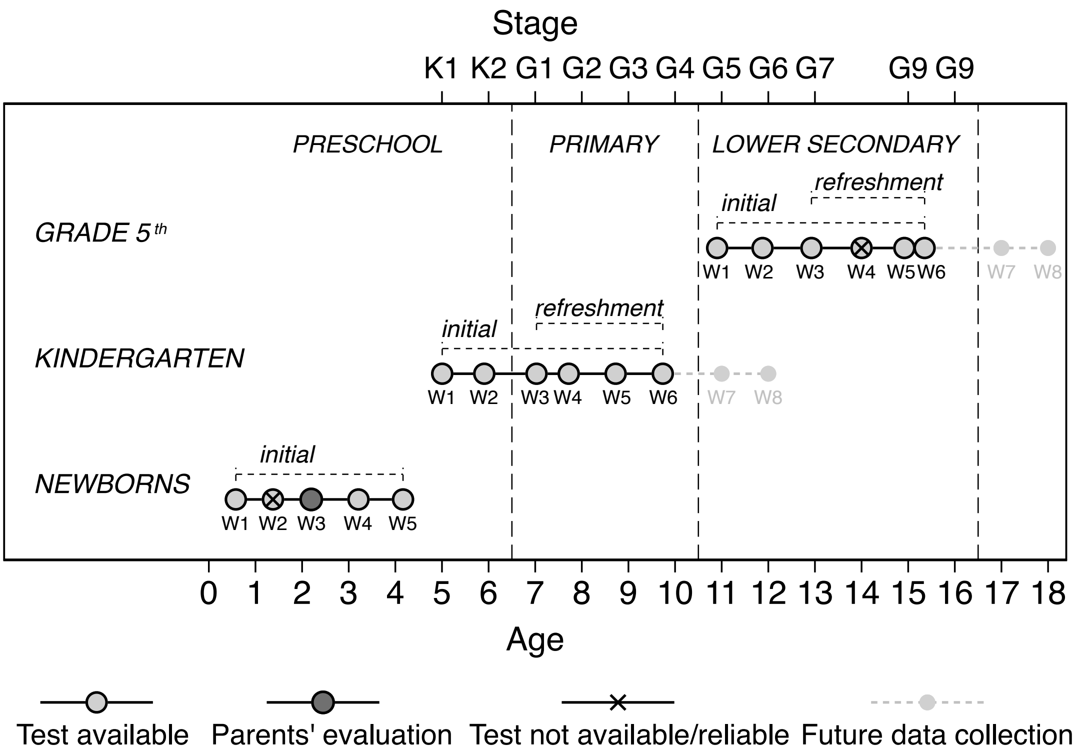


Figure 1 Structure of the linked longitudinal design based on the NEPS data.

2.3.2 Competence assessment (skills)

Our primary interest is on disparities in the cognitive component of children's development. The NEPS offers undoubtedly the wealthiest set of data among the countries included in this report as children are tested on a wide array of cognitive domains including *domain-general* (e.g. basic cognitive skills), *domain-specific* (e.g. vocabulary), and *stage-specific* competencies (e.g. orthography). The analyses included in this chapter rely on a total of 57 competence assessments. A detailed list of all the competence assessments used is in the Appendix (see Appendix 2.2).

We focus on two set of outcomes. First and foremost, we analyse several competence domains related to learning which represent fundamental skills for the acquisition of educational credentials and labour market success: *mathematics, science, vocabulary, reading*, and *orthography*. Moreover, we analyse possible disparities in *basic cognitive skills*, which offer an interesting comparison as they are more biologically rather than environmentally determined. As the NEPS does not consistently test the same domain in each adjacent wave, the analyses are restricted in terms of time points and overall observation window depending on the domain considered. The six selected domains are nonetheless the only domains for which there are at least four measurements available over the entire observation window.

In addition to the six single competence domains, we constructed a *composite measure* of achievement combining all competence tests available in each wave. This composite measure considers the qualitative change in children's development over the early years and offers a synthetic portrait of the evolution of disparities across the overall time window (from birth to adolescence).

2.3.3 Socio-economic background

Information on the family environment and the socio-economic background are gathered *via* a questionnaire administered to parents (mostly mothers). We use two complementary measures of the socio-economic background: *parental education* and *parental income*. Both measures date back to the first participation of target children in the study.

Parental education is measured by the years of education that are necessary to acquire the educational qualification of the parents. We apply the dominance approach and distinguish three broad categories based on highest education level among the parents: ‘low' (12 years of schooling or less), ‘medium' (between 13 and 15 years of schooling) and ‘high' (above 15 years of schooling). In the case of single parents, we only took the educational level of the interviewed parent. Children in the low category are approximately around 15% in our representative cohort samples.

Household income is used as a complementary measure of social background. As the primary interest here is on deprived families, we constructed a threefold classification based on the relative position of the families in the overall distribution of households’ income in Germany. The values used as thresholds for the three categories – low- medium- and high-income – are taken from the German Mikrocensus (in the year 2011) and distinguished the lower 30%, the middle 40%, and top 30% of the population in Germany in 2011. Note that NEPS data are representative of a positively selected share of the population in Germany (families with children). As a consequence, the thresholds mentioned above identify a lower share of low-income families in our representative cohort samples (around 10%).

In some analyses, we use a parental education (highest years of education among the parents) and household income as metric variables in order to cope with the small sample sizes at hand.

2.3.4 Migration background

Information on migration background is also gathered from the parent questionnaire and date back to the first participation of the children in the study. At the baseline level, we distinguish first- and second-generation migrants from children with no migration background (children born in Germany whose parents were also born in Germany). This distinction is used to analyse the *migrant-native* gap in educational achievements.[2]

In addition to the overall migrant-native gap, we also provide a more nuanced portrait of education inequalities by focusing on two *specific groups of migrants*: Turkish and Russians. As anticipated, limitations of the data (lack of details on the country of origin) do not allow us to distinguish other groups of migrants, such as Maghrebian and Roma.

[2] For the sake of readability, we refer to ‘natives’ only in case of children with no migration background. We are well aware that, in principle, also second-generation migrants are ‘natives’ in the sense that they are born in Germany.

2.4 Methods

2.4.1 Construction of standardised scores

We adopt a relative approach to the measurement of achievement inequalities (see Chapter 1). Test scores from the six competence domains and the composite measure are standardised in each wave to reflect the relative position of children in the distribution of achievement at each time point. The composite measure of achievement reflects the simple average of all standardised scores available in each wave for a total of 57 tests over the entire observation window.

A general problem in the construction of standardised scores is that – although tested in the same wave – children may vary slightly in their age. As age influences itself the process of cognitive development, systematic age-variations by SES and migration background may confound our measurement of achievement inequalities. This confounding may be particularly problematic in the institutional samples of SC2 and SC3, which are representative of children attending kindergarten and Grade 5, respectively. The problem is absent in SC1 insofar as it is representative of a birth cohort of children. Similar to Bradbury and colleagues (2015), we addressed the issue by cleaning standardised scores from the confounding effect of age-variations *via* residualisation. Substantively, our procedure ensures that standardised scores within each wave capture solely differences among children at the same developmental stage. Weights are used throughout the construction of standardised scores to ensure that they reflect variations in the target populations and not in our samples. The Appendix (see Appendix 2.3) provides a detailed account of all procedures involved in the construction of the standardised scores.

2.4.2 Describing the extent and evolution of achievement gaps

The description of the evolution of achievement gaps relies on points averages computed on the unbalanced panel of children. More specifically, we compare the average relative position of children with different SES and migration background over the extended observation window. Points averages (conditional means) are calculated separately in each wave and competence domain using Ordinary Least Squares regression models.

As regards SES, we compute the average relative position among groups of children with high, medium, and low educated parents and from families with high, medium and low household income, respectively. The average standardised scores by SES are calculated controlling for possible variations in the composition of high, medium, and low SES families by migration background. Intuitively, this procedure ensures that SES differences in the scores reflect only inequality mechanisms related to the socio-economic status rather than the migration history of the family of origin. The evolution of SES differences is presented by plotting, over time, the average relative position of children from the three SES groups, separately for the six single competence domains and the composite measure of achievement.

As regards migration background, we computed differences between migrants and natives generally, and then natives, Turkish, and Russians specifically. Here, we prefer plotting differences in the average relative position between migrants and natives (and between the specific groups of migrants and natives) directly. Negative values imply migrants' penalties while positive values imply migrants' advantages. Gaps are calculated first ignoring and then considering possible variations in the composition of migrants' groups in terms of SES. The first approach allows us to inspect gross differences between the average relative position of natives and migrants – the *overall* migration effect. The second approach allows us to inspect differences in the scores that only depend on the migration history of the family of origin and not on their SES – the *net* effect of migration background. Intuitively, this second quantity may be interpreted as the achievement gap between migrants and natives hailing from families with the same SES. Given the small sample size in groups of migrants, *net* gaps are inspected using a parsimonious metric variable measuring parental education as the highest years of education of among the parents (results using household income as a metric variable are very similar). The ratio between the *net* and the *total* migration background effects is used as a relative measure of the extent to which SES explains the differences between the overall performance of migrants and natives.[3] As there is one ratio per time point, we computed the average of the ratios across all time points, separately by each competence domain (the average ratios are multiplied by 100 so that they are interpretable on the percentage scale). While offering a synthetic and straightforward interpretation, these ratios are relative measures and are sensitive to the absolute values of the overall gaps. Therefore, small negative or positive *overall* gaps may mirror very high percentage explained by SES although the role of SES may be substantially minimal. As there are substantial differences in the extent and evolution of achievement gaps by competence domain in the case of migration background, we avoid analysing the composite measure and focus on the single domains only.

Weights are used throughout the analyses. Details on the specifications of the models used can be found in the Appendix (see Appendix 2.4).

2.4.3 Explaining school inequalities by preschool inequalities

Explaining school inequalities through preschool inequalities requires observing how the same children perform in preschool and subsequently over the school years. Therefore, this second set of analyses requires focusing on a balanced panel of children only. We limit the analyses to the original sample of SC2 (interviewed in the first wave), which are the only children assessed in preschool (the refreshment sample starts when children are already enrolled in primary school). Only two domains are assessed in the first wave of SC2 and at least two additional times over the educational career: vocabulary and science.

We choose to focus on vocabulary only for several reasons. First, vocabulary is one of the most critical dimensions of cognitive development in the early years as it favours the acquisition of skills in other competence domains. Second, the focus on vocabulary increases the comparability with the previous findings as vocabulary is commonly analysed in the previous literature relying on both cross-sectional and longitudinal designs. Third, vocabulary is the domain in which SES inequalities are among the largest and the migrant-native gap follow a fascinating pattern of evolution (as we will show later). Children of SC2 took vocabulary tests in three instances: the first year of kindergarten (age 5), and the first and third year of primary education

[3] Assuming that migrants perform worse than natives overall, a ratio below 1 – for example .7 – indicates that SES does not fully explain the migrant-native gap – only the 30% in the example (1–[net/tot]). A ratio above 1, for example, 1.2, implies that the migrants' penalty is even reversed when comparing families with similar SES and that now migrants experience an advantage which is the 20% of the original penalty. A ratio of 1 implies that SES fully explains the overall migrant-native gap.

(age 7 and 9, respectively).

The sample size of the balanced panel is quite limited: around 400 subjects. This sample size puts some limitations on the analyses. First and foremost, it is not feasible comparing the specific groups of migrants (Turkish and Russians) with natives, although we can still analyse the gap between natives and migrants *tout court*. Second, we are forced to use metric measures of parental education and household income for reasons of parsimony. Third, the analyses must be conceived as simulations rather than descriptions since the uncertainty around the estimates is considerable.[4]

Similar to the description of the evolution of achievement gaps, we rely on conditional means. However, as RQ2 requires conditioning school achievements (age 7 and 9) on preschool achievements (age 5), OLS is not an adequate tool and would result in biased conclusions. The problem occurs since children may perform above or below their actual ability in preschool, for example, because they had a good or bad day or only due to luck. However, lucky children are not likely to be as fortunate in subsequent assessments, and their scores would regress on their actual ability level. Insofar as there are differences in the average preschool achievement by groups, say by SES, children deviating from the average group performance would regress to their group-average in subsequent assessments, thus biasing the interpretation of group-level trajectories over time. The problem is known as *regression to the mean* and is solved adopting an instrumental variable approach (see Bradbury et al. 2015; Jerrim & Vignoles 2013). Intuitively, our strategy undertakes regression to the mean by exploiting the correlation between preschool tests in vocabulary and other competence domains tested in a different day. Details about all procedures involved are in the Appendix (see Appendix 2.4).

We investigate conditional achievement gaps by SES (both parental education and household income) and migration background (both *overall* and *net* gaps, as defined in the previous section). Similar to Bradbury and colleagues (2015), we organise the analyses in three steps. In the first step, we plot the average relative position of children with different social and background over the three time points available. In addition to these observed trajectories, we simulate the school trajectories we would find if preschool scores were the only predictors of later scores and social and migration background had no role for achievement inequality beyond preschool.[5] If these conditions hold, then the observed and the simulated trajectories will overlap entirely. Hence, differences between observed and simulated trajectories indirectly suggest the relative importance of preschool inequalities and the additional role played by social and migration background over the school career for the explanation of disparities observed in primary school.

In a second step, we estimate the proportion of the achievement gaps in the first (age 7) and third year (age 9) of primary education which is explained by preschool inequalities (age 5). This decomposition relies on a simple mediation analysis in which preschool achievements mediate the total effect of SES/migration on primary school achievements. We use the ratio between the direct and the total effect as a relative measure of the extent to which inequalities over primary school are explained by the direct role of SES/migration over schooling (say *x*) and, more importantly, by preschool inequalities (say *1–x*). The ratios are multiplied by 100 to allow for

[4] For this reason, we do not plot confidence intervals in the figures.

[5] The simulated trajectories are obtained estimating the school performances that children with a preschool performance equal to average high, medium and low SES child and the average native/migrant would have in a world where we were agnostic about their actual social and migration background (i.e., without including SES and migration background in the model).

a straightforward interpretation on the percentage scale. Worth noting is that these ratios may assume negative values if the direct effect of SES/migration has opposite sign compared to the total.

In the final step, we simulate the trajectories we would observe if the influence of preschool scores on school achievements would differ among children with a different social and migration background. This simulation tentatively suggests whether children starting with equal preschool performances develop differently depending on their social and migration background. In other words, these final analyses elucidate whether the additional role of social and migration background over the school career concentrates among children at the top, the middle, or the bottom of the distribution of achievements in preschool.

As usual, we use weights (balanced) throughout the analyses. Details about the all the estimated empirical models are in the Appendix (see Appendix 2.4).

2.5 Empirical results

2.5.1 Extent and evolution of achievement gaps (RQ1)

2.5.1.1 Inequalities by socio-economic status

When do SES gaps in achievements arise and how they evolve over the educational career?

Figure 2 presents the evolution of the SES gaps in achievement measured by parental education (panel A) and household income (panel B) broken by the six single competence domains. The points on the graph represent the average relative position of children with high, medium and low SES in each domain and time points available over the early life course. Overall, the figure shows noticeable and persistent – or even increasing – SES gaps in achievement. SES gaps seem to emerge early in the life course: low SES children score systematically lower compared to high SES children even before the formal enrolment into the school system (age 7 in Germany). Panel A shows that the early gaps between children with high and low educated parents tend to increase moving from the preschool period throughout primary and lower secondary education in all competence domains.

When measuring SES by household income (panel B), instead, this tendency towards increasing gaps is less pronounced and a picture of stability prevails. Moreover, gaps measured by household income are somewhat lower and more fluctuating (due to fluctuations in the low-income group) compared to gaps measured by parental education. While these subtle differences may reflect the relative importance of different parental resources, they may also reflect differences in the quality of the two measures. As we suggested in the method section, parental education is a more suitable measure in our context as it is a more stable measure of social background compared to parental income (especially among young families). Therefore, in the remainder of the chapter, we will mainly stress results related to parental education while still reporting results for household income.

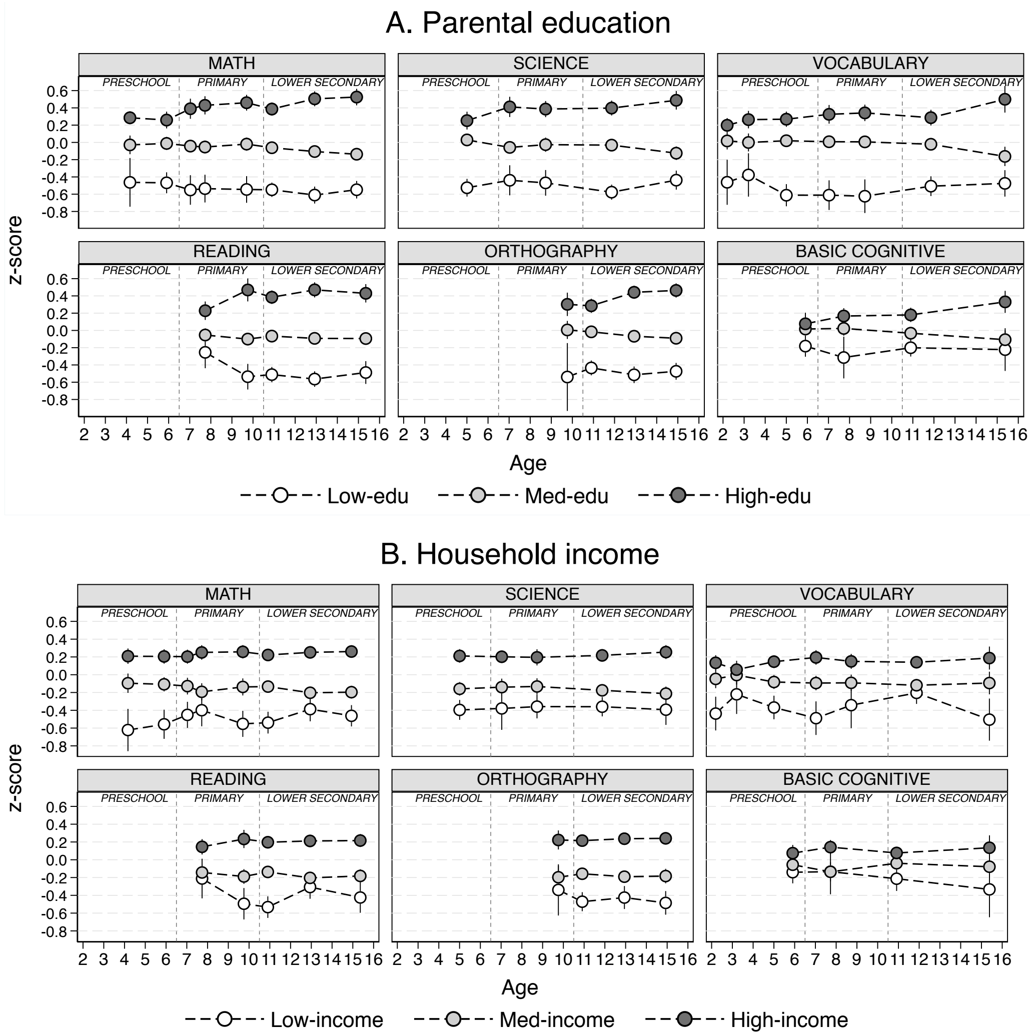


Figure 2 Domain specific z-scores of children by parental education (A) and household income (B).

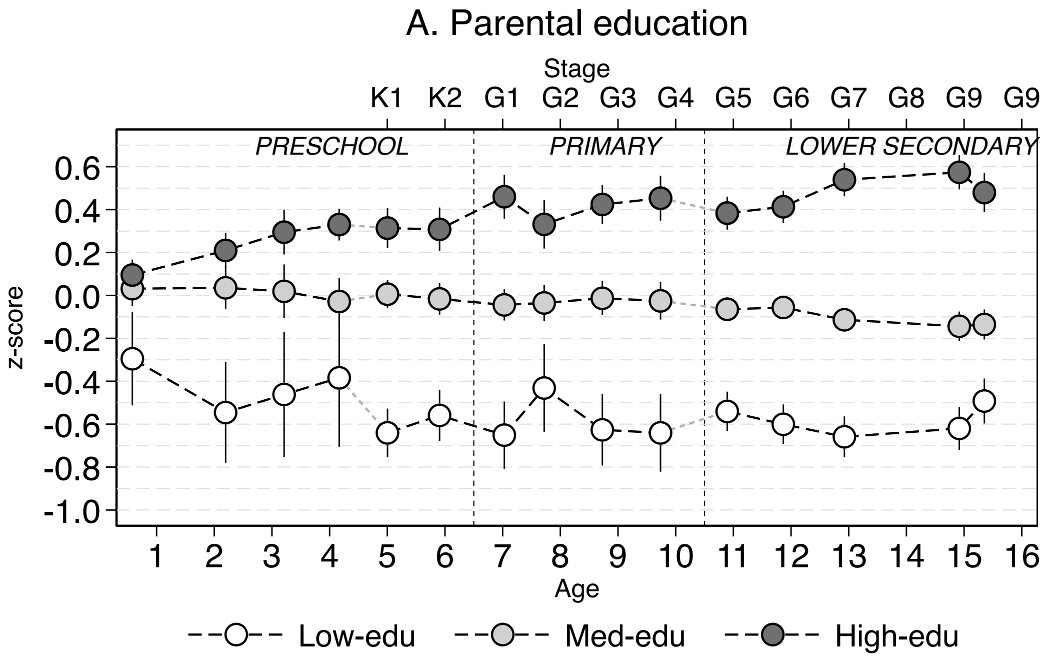
Notes: Predictions based on life stage-specific regression models. 95% confidence intervals for predictions shown.

Although characterised by the same pattern of persistent or even increasing inequality over the educational career, Figure 2 also shows also some interesting heterogeneities in the size of gaps across competence domains. Overall, the difference between high and low SES children is substantially higher (almost doubled) in the five competence domains related to learning – math, science, vocabulary, reading, and orthography – compared to basic cognitive skills – the only outcome associated with fluid intelligence. This result holds both when measuring SES by parental education or household income (in the case of household income, the uncertainty around the estimates questions whether there are differences at all).

Figure 3 complements the analysis of the single competence domains by focusing on the composite measure of achievement. The interpretation of the points on the graph is the same as before: they represent the average relative position of children with high, medium, and low parental education (panel A) or household income (panel B) in the distribution of the composite measure in each time point. Compared to the analyses of the single domains, the composite measure offers a straightforward summary of inequalities and allow us to evaluate the evolution of these inequalities over the entire observation window – from birth to the end of lower secondary education (age 16).

The analysis of the composite measure confirms the existence of remarkable SES gaps in achievement that settle down very early in the life course. SES differences are even visible at 7 months of age, increase over the early years and settle down before the school entry (age 6). After that point, SES gaps tend to grow at a slower pace. For example, the average difference between children with high and low educated parents (panel A) is .4 SD at 7 months of age, .9 SD in the last year of the preschool period (age 6), 1.1 SD in the end of primary education (age 10), and 1.1 SD in the end of lower secondary schooling (age 15-16).

As for the single domains, inequalities by household income (panel B) are generally less pronounced. However, while being more fluctuating and somewhat more persistent after the enrolment into primary education (age 7), differences by household income follow a similar increasing pattern over the preschool period (the gap high-low gap increases from .3 SD at age 6 months to .75 SD at the end of preschool and then fluctuates around this value).



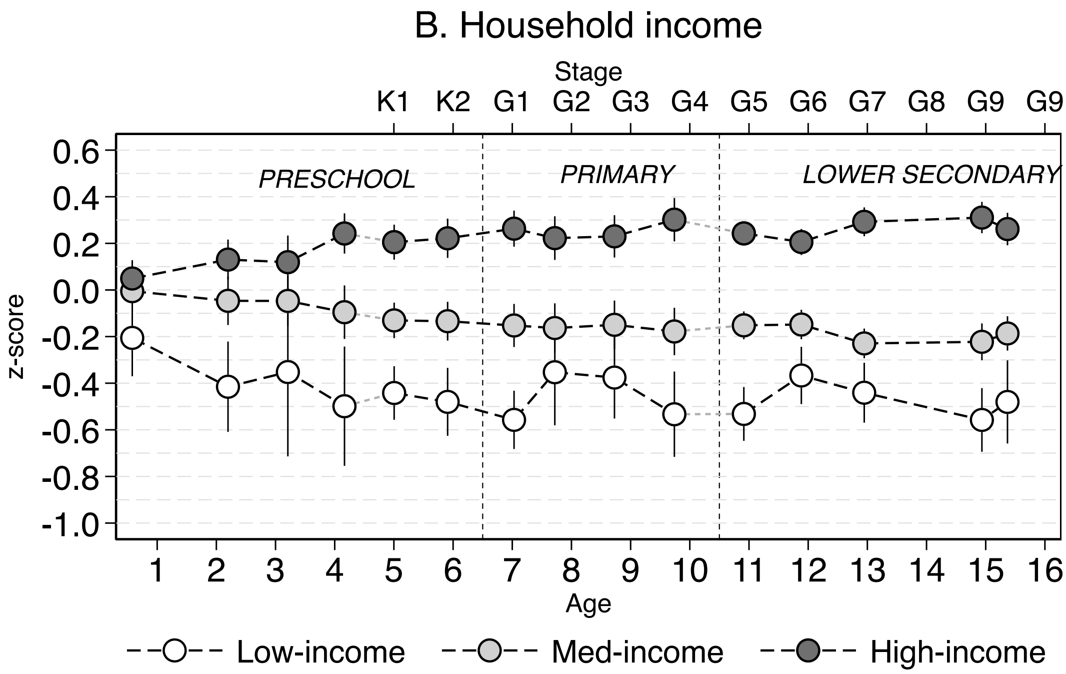


Figure 3 Composite z-scores of children by parental education (A) and household income (B).

Notes: Predictions based on life stage-specific regression models. 95% confidence intervals for predictions shown. Stage: K = Kindergarten, G = Grade level in school. Long-dashed black lines connect data within the same NEPS cohort.

2.5.1.2 Inequalities by migration background

THE MIGRANT-NATIVE GAP

When do migrant-native gaps in achievement arise? Do gaps tend to close or reinforce over the educational career?

Figure 4 shows the evolution of migrant-native gaps in achievement over the early life course broken by competence domain. We present differences in the average performances of migrants and natives (interpretable as average differences between the two groups). Positive values indicate that migrants perform better than natives, while negative values indicate that migrants perform worse than natives. Black dots represent the overall average differences between migrants and natives. Grey dots represent the residual differences once composition of the two groups in term of SES is controlled for, that is, the portion of the migrant-native gap not attributable to typical variations in the social standing of migrants' and natives' families. At the bottom of each subgraph, we report the average percentage of the overall gap explained by SES across all time points available in each domain.

Overall, the figure clearly shows that natives outperform migrants in most of the competence domains considered. By looking at the overall gaps (black dots), it is striking that migrants' penalties are particularly pronounced in the case of vocabulary and, although to a lower extent, science and math. The migrant-native gap, instead, is significantly smaller for reading and orthography, while being almost absent for basic cognitive skills. These differences across domains are understandable since basic cognitive skills are more biologically determined, while vocabulary, in particular, may be more sensitive to learning environments and specifically to the language spoken at home.

As regards the evolution of the overall disparities over the life course, patterns are quite mixed depending on the domain considered. While the migrant-native gap in science seems to decrease moving from preschool to primary school and to remain stable afterwards, estimates for math are more fluctuating and suggest substantial stability of the gap throughout the early life course. Instead, the (lower) migrants' penalty in reading shows a slight tendency towards an increase moving from the early years of primary education to the end of lower secondary education. Conversely, the limited migrant-native gap in orthography seems to vanish moving from the end of primary to the end of lower secondary schooling. Finally, gaps in vocabulary show a peculiar pattern. Migrants' penalty is significantly lower in the early years of life (2–3 years of age), increase dramatically during the preschool period (in coincidence with the entry into kindergarten), and then tends to decrease significantly moving to the primary and especially the lower secondary school years. A possible explanation for this U-shaped pattern is that children's vocabulary is less sensitive to the family environment and the language spoken at home when they are about to pronounce the first words, while the role of the family environment kicks in tremendously as children progress in their verbal abilities around kindergarten age. After that period, the exposure to schooling, for example through interactions with peers and teachers, may be responsible for the attenuation of vocabulary gaps. Migrants' vocabulary may benefit more from schooling compared to natives' insofar as the counterfactual situation at home – the extent and quality of exposure to native language – is worse for the former than the latter.

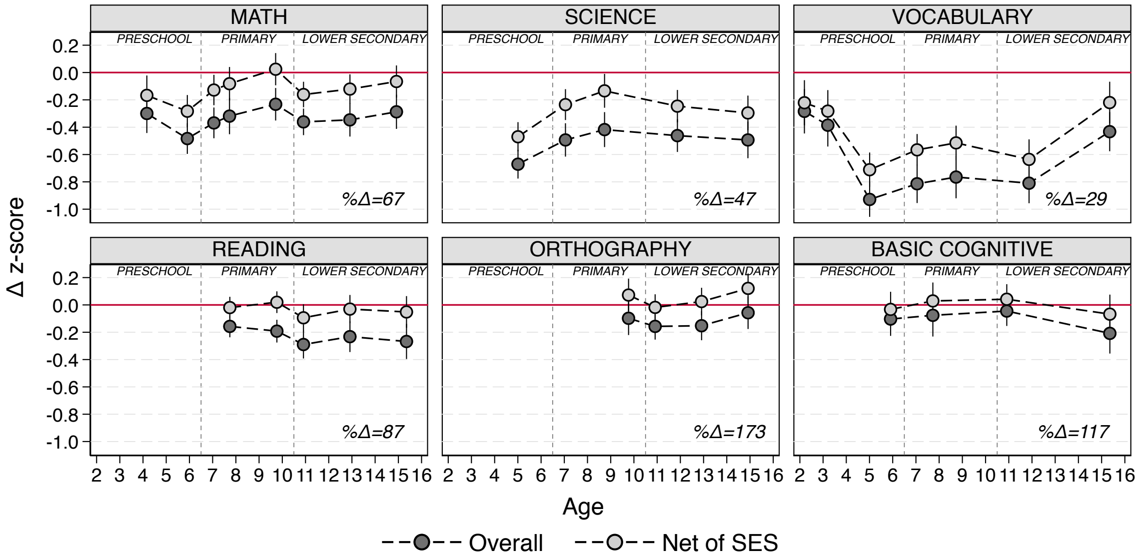


Figure 4 Total and net (SES) migrant-native gaps in various competence domains.

Notes: Predictions based on life stage-specific regression models. 95% confidence intervals shown. %Δ is the average proportion of the overall migrant-native gap explained by SES across the various time points in each domain.

However, to what extent are these overall gaps attributable to SES differences among migrant and native families?

Figure 4 (grey dots) shows that migrant-native gaps are significantly reduced or in some instances even absent when we consider the composition of families by SES. This result suggests that a considerable proportion of the overall migrant penalty can be explained by the lower SES of migrants' families compared to natives', which is itself responsible for lower achievements (as shown in Section 2.5.1.1). Family SES seems fully responsible for the small migration penalty in reading and orthography: migrants and natives children born into families with a similar social standing seem to perform equally in these domains. By looking at the average proportion explained, we can conclude that family SES accounts on average for around the 87% and 100% of the migrant's penalty in reading and orthography, respectively (values of delta above 100 suggest that the gap reverses, meaning that migrants perform better than natives).

Children with a migration background keep performing worse compared to natives with a similar SES in science and especially vocabulary, even if the overall gap reduces by approximately 47% and 29%, respectively. Therefore, there are differences in the proficiency in science and vocabulary that cannot be solely traced back to disadvantages connected to the generally lower social standing of migrant families compared to natives. Finally, the results for math are more mixed and do not provide clear-cut evidence: while it is clear that SES explains the greater proportion of the overall migrants' penalty – around the 67% – it is not clear whether the remaining gap is due to random fluctuations or is a substantial feature of the comparison among the two groups.

Interestingly enough, while explaining the overall migrant-native gaps to a considerable extent, family SES does modify the patterns of evolution of achievement gaps over time.

FOCUS ON MIGRATION GROUPS: TURKISH AND RUSSIANS

Are the extent and the evolution of the migration gaps similar or diverging across different groups of migrants?

Figure 5 shows once again the evolution of migrant-native gaps for two specific groups: Turkish and Russians. The points on the graph reflect differences in the average performances of Turkish (black dots) and Russians (grey dots) compared to natives. Panel A reports overall gaps, while panel B reports the residual gaps once controlled for SES (and the percentage of the overall gap explained by SES). The interpretation of overall and residual gaps is the same as in Figure 4.

Panel A shows that there are differences in the extent of the overall migrant-native gap between Turkish and Russians. Although both groups suffer a penalty, children with a Russian background perform more similarly to natives compared to children with a Turkish origin. These differences are astounding if we consider the five learning domains (excluding basic cognitive skills): while the Russian-native gap ranges from 0 to a maximum of .8 SD, the average performance of Turkish is between .4 and 2 SD lower than natives.

There are clear variations by competence domains, however. First of all, in line with the results presented earlier, neither Turkish nor Russian perform differently in the test aimed at capturing basic cognitive abilities. However, while Russians perform equally to natives in orthography and only slightly worse in reading (around –.2 SD) and math (around –.2/–.3 SD), the average difference between Turkish and natives is by far higher in all three domains (around –.5 SD in orthography, and about –.4/–.8 SD in reading and math). Even in the domains in which also Russians lag behind natives to a considerable extent – such science (–.3/–.8 SD) and vocabulary (–.4/–.8 SD) – the average performance of Turkish is by far the lowest (around –.8/–1.2 SD in science and even up to –2 SD in vocabulary).

Finally, there are also interesting differences in the evolution achievement gaps across the two groups of migrants, although the results are quite mixed depending on the domain considered. While the Russian-native gap in science seems to decrease slightly moving from preschool to the end of secondary education, Turkish seem to lose ground slightly compared to both natives and Russians. A similar pattern regards reading abilities: the Russian-native penalty is quite stable, while the Turkish-native penalty is opening up when transitioning from primary to the end of lower secondary education. In the case of orthography, instead, the Turkish-native gap is quite constant, while Russians seem to fully recover from the small gap observable at the end of primary education and perform similarly to natives starting from the first year of secondary school. Finally, patterns for math and vocabulary does not vary across the two groups: differences in math are fluctuating and suggest stability, whereas in the case of vocabulary we observe the usual U-shaped pattern (increasing and then decreasing gaps).

All in all, children with a Turkish background are worse-off compared to natives and even compared to the other most prevalent group of migrants in Germany (Russians). What is more, while the Russian-native penalty tends to reduce over the educational career, the Turkish-native penalty seems to even open up slightly, at least in some domains.

To what extent are differences in the performance of Turkish, Russians and natives attributable to differences in the SES of their families? Are the lower performances of Turkish and Russian compared to natives a story of socio-economic deprivation or rather a story of cultural *differences?*

Panel B in Figure 5 sheds some light on these issues by showing the residual migrant-native gap once controlled for the composition of Turkish, Russian, and native families in terms of SES. The overall Turkish- and Russians-native gaps reduce when comparing children stemming from families with similar SES. This result supports the idea that SES is an essential driver of both Turkish and Russians penalties in achievement. There are differences between Russians and Turkish, however. Once SES conditions between families of origin are equalised, there are virtually no differences in the performance of Russians and native children in math and reading, while the gap in orthography is even reversed (Russians perform even better than native). Instead, there are still some residual gaps in vocabulary and, to a lower extent, science, although in both cases gaps seem to vanish moving towards the end of lower secondary education (in the case of vocabulary we observe the usual U-shaped pattern). We can tentatively argue that SES explains the 100% of the Russians' penalties in math, reading and orthography, and approximately the 70% and the 40% of penalties in science and vocabulary, respectively.

However, SES does not fully explain gaps between Turkish and native children: even when comparing children stemming from families with similar SES, Turkish children still perform below the level of natives in all domains related to learning. The Turkish-native gap in science is still present in the early years (around age 4–5) and tends to increase over the educational career slightly. Although absent in primary education, net gaps in reading and orthography tend to show up at later educational stages. The math gap is instead more fluctuating but still well pronounced. Finally, a massive divide in the mastery of German vocabulary remains between Turkish and natives (again, we can observe a U-shaped pattern for vocabulary over the educational career). All in all, we can tentatively conclude that SES explain less than a half of the overall Turkish-native gaps in the competence domains related to learning: around the 25% in vocabulary, and around the 40–55% in math, science, reading, and orthography.

Concluding, we can argue that differences in the performances of Russians and native children are mainly a story of socio-economic deprivation, while differences between Turkish and native children cannot be solely traced back to the low socio-economic conditions that generally characterises Turkish families in Germany.

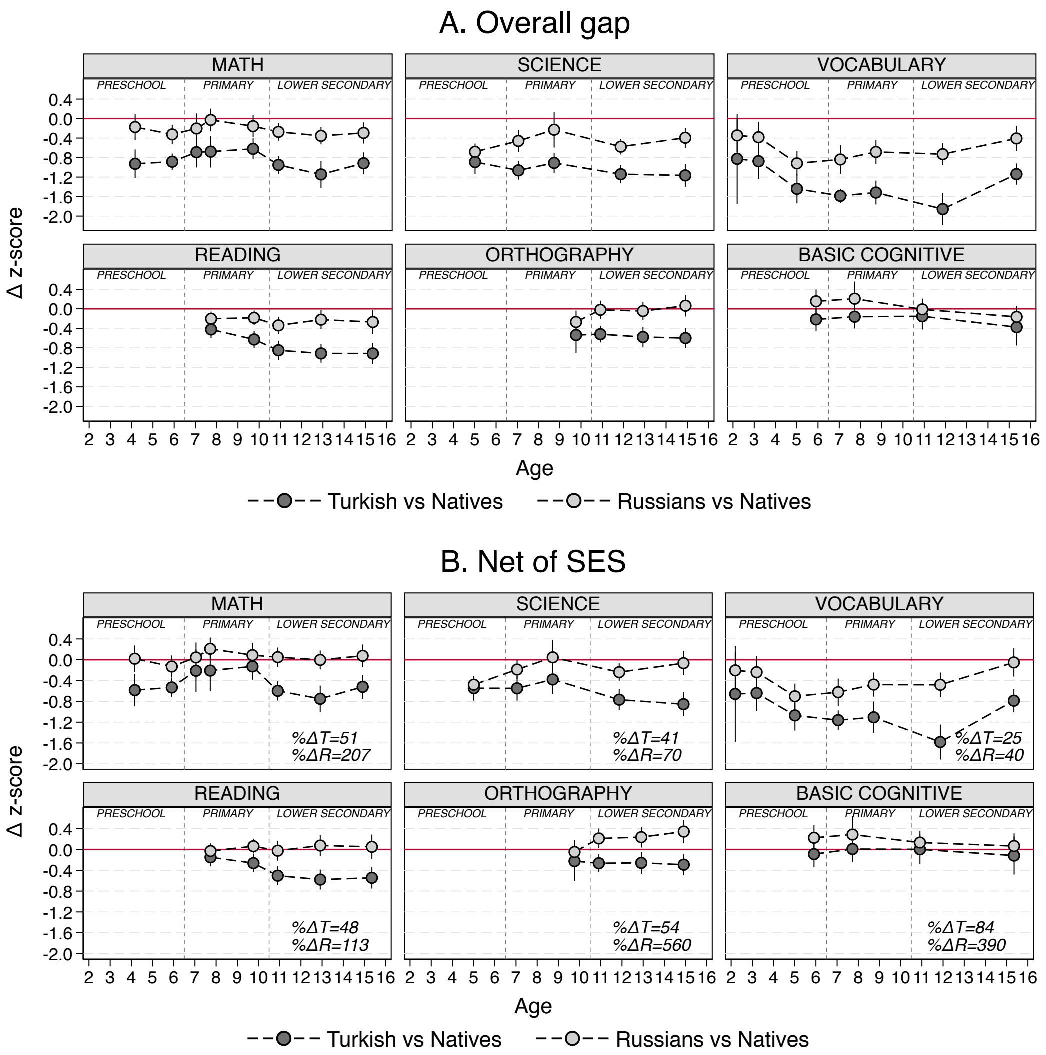


Figure 5 Total (A) and net (B) Turkish- and Russian-native gaps in various competence domains.

Notes: Predictions based on life stage-specific regression models. 95% confidence intervals shown. %Δ is the average proportion of overall Turkish and Russian-native gaps explained by SES across the various time points in each domain.

2.5.2 Explaining school inequalities by preschool inequalities (RQ2)

2.5.2.1 Inequalities by socio-economic status

Do preschool inequalities explain SES gaps in educational achievement in school?

Figure 6 shows the evolution of achievement gaps in the domain of vocabulary moving from the preschool period (age 5) to the first (age 7) and the third year (age 9) of primary education. Dots with black outlines indicate the observed average proficiency of high-, medium-, and low SES children (left and right panels measure SES in terms of parental education and household income, respectively). Dots with shadowed outlines simulate the trajectories that children with preschool performances (age 5) equal to the average high-, medium- and low SES child would have if preschool scores were the only predictors of later scores and SES had no role for achievement inequality beyond preschool.

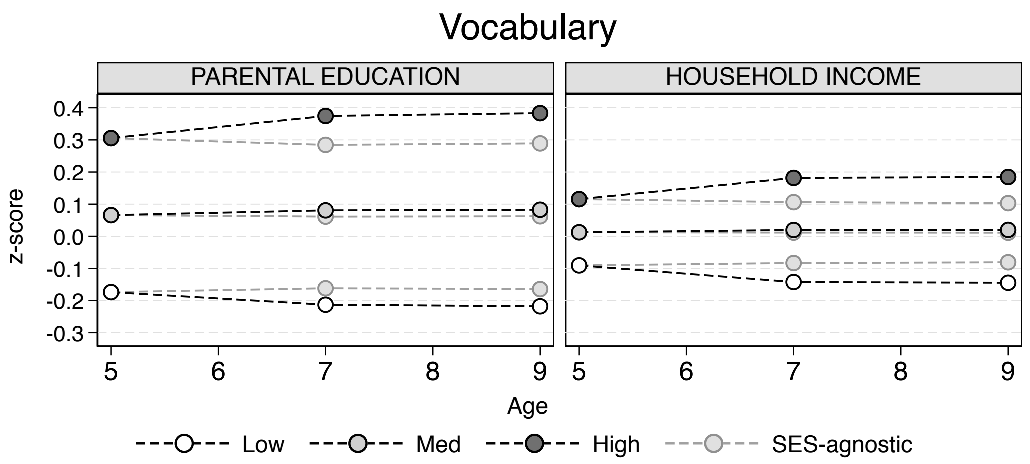


Figure 6 Observed and simulated (SES-agnostic) trajectories moving from preschool to the end of primary schooling by parental education (left panel) and household income (right panel).

Notes: Predictions based on life stage-specific regression models. Simulated trajectories reflect the z-scores we would observe in primary school if the preschool z-scores were the only predictors of later z-scores and SES had no effect beyond preschool (primary school z-scores predicted based on the average high- medium- and low-SES performances in preschool).

The observed trajectories show that SES inequality in vocabulary is already set in the preschool period and then increases only slightly moving from preschool to the first and third year of primary education. Interestingly, observed trajectories also show that inequality by parental education is generally higher compared to inequality by household income (although the two measures are not strictly comparable). These findings, which only rely on a balanced panel of children, are generally in line with what is found looking at the unbalanced panel (see Figure 2).[6] However, the figure clearly shows that the observed trajectories are not entirely in line with the trajectories we would observe if SES had no role beyond preschool and preschool performance was the only predictor of later performances. The simulated trajectories, instead, suggest that SES inequality would be lower if SES played no role beyond preschool: rather than increasing slightly over time, SES inequalities would be perfectly stable under such conditions. Although differences between observed and simulated trajectories are not dramatic, these differences suggest SES may play a role beyond preschool and that SES gaps in school cannot be solely traced back to SES inequality already present in the preschool period.

[6] The only exception is the evolution of inequalities by household income, which is more fluctuating rather than slightly increasing when looking at the results from the unbalanced sample. However, we cannot formally compare the results of Figure 2 and Figure 6 as they rely on different measurements of parental education and household income (groups vs metric variables).

Table 2 Total and direct (net of preschool z-scores) SES effect in the first (age 7) and the third (age 9) years of primary education. Separately for parental education and household income.

	PARENTAL EDUCATION				HOUSEHOLD INCOME			
	AGE 7		AGE 9		AGE 7		AGE 9	
	TOTAL	DIRECT	TOTAL	DIRECT	TOTAL	DIRECT	TOTAL	DIRECT
SES	0.147***	0.042	0.150***	0.044	0.065	0.028	0.066	0.03
	(0.027)	(0.033)	(0.025)	(0.029)	(0.046)	(0.026)	(0.037)	(0.023)
SCORE AGE 5		0.872***		0.884***		0.904***		0.876***
		(0.088)		(0.103)		(0.09)		(0.096)
% PRESCHOOL		71		70		58		55
% ADD SES		29		30		42		45
N	400	400	400	400	398	398	398	398

Notes: Parental education = 1-year units. Household income = 1000 Euro units. All models control for migration background. Cluster robust standard errors in parentheses. Significance: * p<0.05; ** p<0.01; *** p<0.001.

However, to what extent are SES inequalities in primary school attributable to preschool differences?

Table 2 shows an estimation of the proportion of SES gaps in the first (age 7) and third (age 9) year of primary education which is explained by preschool inequalities (age5) and the additional role of SES unfolding beyond preschool, separately for parental education (left panel) and household income (right panel). Once controlled for typical differences in the preschool achievements of high, medium, and low SES children, a residual (direct) positive SES effect on primary school scores remains. Although not statistically significant – possibly due to the small sample size of the balanced sample – this direct SES effect suggests that SES differences in preschool scores alone may not be able to explain SES differences in primary school fully.

When looking at parental education, it seems that preschool differences are responsible for around 70% of SES gaps we observe in the first and the third year of primary school. While this estimate may be considered a lower bound (due to the non-statistical significance of the direct SES effect), it may be the case that the remaining 30% of primary school gaps is attributable to the additional role that SES plays over the school careers. Interestingly, the relative importance of preschool achievement is lower when it comes to household income: around the 55–60% of primary school gaps may be attributable to preschool differences. However, the uncertainty around the estimates in the case of household income is high and prevent us from any further speculation. All in all, Table 2 clearly shows that most of SES differences in vocabulary during the primary school years can be explained by proficiency differences that are already well settled in kindergarten. There may still be some room for the role of SES on top of preschool differences, however. The next natural question is whether the additional role played by SES over the primary school years is or not dependent on the achievement that children have shown in preschool.

Does SES compensate for a lousy start in kindergarten? Or does SES offer an advantage irrespective of the achievement level in the early years?

Figure 7 answers these questions via simulation of the trajectories high, medium, and low SES children would follow when staring at the top, the middle and the bottom of the achievement distribution in preschool. As said, this approach must be conceived as a simulation since the limited sample size results in strong uncertainty around the estimates. Moreover, we have to bear in mind that the size of the additional SES effect is comparably small compared to the influence of preschool differences. Still, Figure 7 shows some intriguing results. The role of SES over the primary school seems to concentrate at the bottom of the distribution of achievement in preschool: when starting low, high SES children seem to fare better over time compared to low SES children performing similarly in kindergarten. However, children from different SES in the middle and the top of achievement distribution in preschool seem to perform similarly over the primary school years. Worth noting is also that, although partially climbing up the ladder over the primary school years, high SES children performing low in preschool are far from fully recovering from the initial penalty compared to high SES children starting in the middle or at the bottom of the distribution of achievements.

These tentative conclusions hold both when analysing SES in terms of parental education and household income. The reasons behind this compensation effect remain unknown, however. On the one hand, it may be that high SES parents deploy efforts to compensate for a child's bad start, for example hiring private teachers, overstimulating the child at home, and/or taking advantage of primary school differentiation in terms of curriculum, teacher quality, and school composition. On the other hand, it may also be that high SES children performing low in preschool are late bloomers regarding cognitive development.

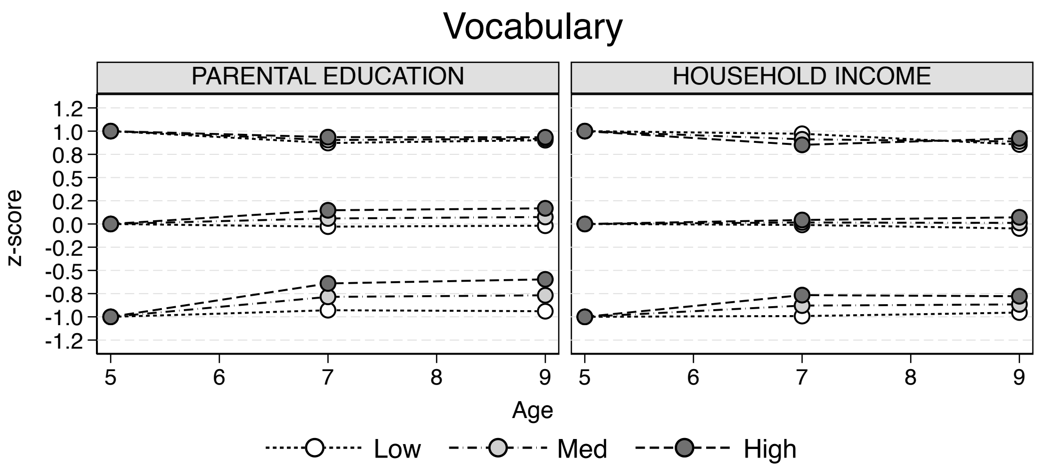


Figure 7 Simulated trajectories moving from preschool to the end of primary schooling of children from different parental education (left panel) and household income (right panel) but starting with the same preschool z-score.

Notes: Predictions based on life stage-specific regression models. Simulated trajectories reflect the z-scores we would observe in primary school if the additional SES effect beyond preschool depended on preschool z-scores (primary school z-scores predicted for children from different SES performing 1, 0 and –1 SD above/below the average preschool performance)

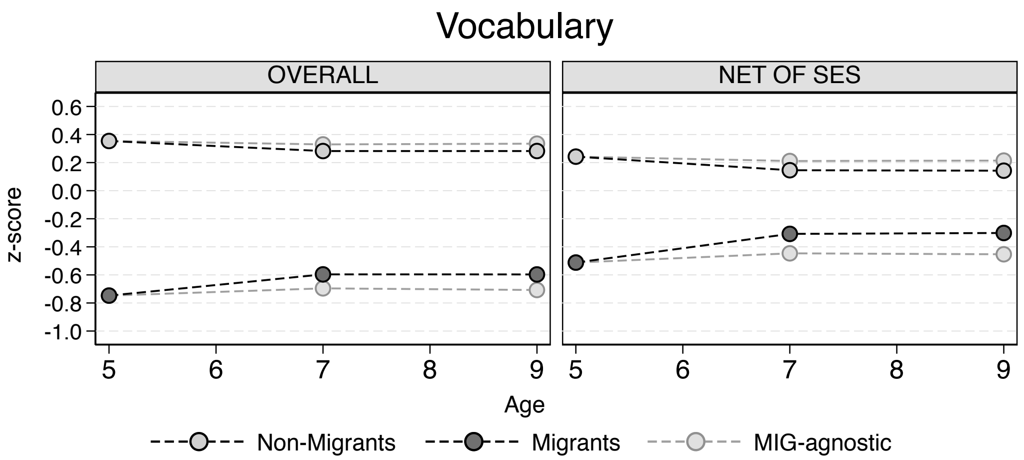


Figure 8 Observed and simulated (migration-agnostic) trajectories moving from preschool to the end of primary schooling by migration background: overall (left panel) and net (right panel) migrant-native gaps.

Notes: Predictions based on life stage-specific regression models. Simulated trajectories reflect the z-scores we would observe in primary school if the preschool z-scores were the only predictors of later z-scores and in absence of additional migration effect beyond preschool (primary school z-scores predicted based on the preschool performance of the average migrant and native child).

2.5.2.2 Inequalities by migration background

Do preschool inequalities explain migrant-native gaps in educational achievement in the school years?

Figure 8 shows the evolution of the migrant-native vocabulary gap from preschool (age 5) to the first (age 7) and the third (age 9) year of primary education. As in the previous section, we only focus on a balanced panel of children. Dots with black outlines indicate the observed average proficiency of migrants and natives. Dots with shadowed outlines simulate the trajectories that children with preschool performances (age 5) equal to the average migrant and the average native would have if preschool scores were the only predictors of primary school scores and migration background played no role for achievement inequality beyond preschool. Left and right panels show the results both for the overall and the net migrant-native gap, respectively.

The observed trajectories confirm some of the findings obtained using the unbalanced panel (see Figure 5): migrants perform on average worse in vocabulary than natives in kindergarten, the migrants' penalty decreases moving throughout primary education, and SES accounts for a significant proportion of the overall-migrant native gap. However, it seems that the simulated trajectories do not entirely overlap with the observed trajectories. This result indicates that achievements in kindergarten are not the only predictor of achievements during school and that migration background may play a role beyond preschool. The simulated trajectories surprisingly suggest that the migrant-native gaps in primary education would be slightly larger if migration background did not influence preschool. This result is surprising as it indirectly indicates that, despite the considerable penalty of migrants in kindergarten, migration background may indeed operate towards a compensation rather than an accumulation of preschool inequalities in achievements. Worth noting is also that this result holds for both the overall and the net migrant-native gap, although we need caution as differences between observed and simulated trajectories are not dramatic and the uncertainty around the estimates is considerable.

However, to what extent are the observed penalties of migrants in primary school explained by preschool inequalities?

Differences between observed and simulated trajectories already offer an answer: preschool inequalities entirely explain the migrant penalty in school. However, observed and simulated trajectories do not inform us about the extent to which migration background may even boost performances beyond preschool. Table 3 provides answers to these questions by showing the percentage of the total effect of migration background on achievement in the first (age 7) and the third (age 9) year of primary education that is explained by preschool differences and by the additional role that migration background plays beyond preschool, separately for the overall (left panel) and the net (right panel) migrant-native gap.

Table 3 Total and direct (net of preschool z-scores) overall and net (SES) migrant-native gaps in the first (age 7) and the third (age 9) years of primary education.

	OVERALL				NET OF SES			
	AGE 7		AGE 9		AGE 7		AGE 9	
	TOTAL	DIRECT	TOTAL	DIRECT	TOTAL	DIRECT	TOTAL	DIRECT
MIGRATION (yes)	–0.879***	0.147	–0.879***	0.163	–0.453**	0.204	–0.444*	0.223
	(0.183)	(0.156)	(0.19)	(0.171)	(0.166)	(0.141)	(0.171)	(0.169)
SCORE AGE 5		0.931***		0.946***		0.872***		0.884***
		(0.100)		(0.112)		(0.088)		(0.103)
SES (parental edu)	no	no	no	no	yes	yes	yes	yes
% PRESCHOOL		117		119		145		150
% ADD MIG		–17		–19		–45		–50
N	400	400	400	400	400	400	400	400

Notes: Parental education as metric variable (units = years). Cluster robust standard errors in parentheses. Significance: * p<0.05; ** p<0.01; *** p<0.001.

The table shows that both the overall and the net migrants' penalties in primary education reverse when preschool inequalities are controlled for: as said, this result implies that preschool inequalities explain the 100% of the migrants' penalty during school. What is more, around the 17–19% and the 45–50% of the overall and net migrant-native negative gap in primary school turn to be in favour of migrants once typical differences in preschool achievements between the two groups ruled out. Therefore, it seems that migrants may even outperform natives in primary school when starting with a comparable level of early achievement. Still, we need to interpret with caution this latter finding as there is considerable uncertainty around the estimation of migrants' advantages when preschool differences are equalised. All in all, based on Table 3, we can conclude that preschool inequalities fully explain migrants' penalties in primary school, while we can only tentatively suggest that migration background is compensating for preschool inequalities along the school career. Finally, we ask whether the possible migrants' advantage over the school years distributes equally among high and low achievers in preschool.

Does migration background compensate for the weak performance of migrants in preschool? Or does migration background boost the achievements of migrants performing already well in the early years?

In Figure 8, we simulated the relative differences in the primary school achievements of migrants and native scoring at the top, the middle, and the bottom of the distribution of vocabulary scores in preschool. Trajectories for the overall and net migrant-native gaps are shown in the left and right panel, respectively. As in the previous paragraph, this step of the analyses presents a simulation rather than a description due to the low sample size and the uncertainty about the existence of a migrants' premium on top of preschool differences. The figure shows that the migrants' advantage on top of preschool differences mostly concentrates among medium and high achievers in preschool. Therefore, migrants who perform high in kindergarten seem to increasingly gain ground over the primary school years compared to natives who perform equally well in kindergarten. This boosting effect is somewhat lower in the middle of the distribution of preschool achievement. Finally, migrants and natives performing equally low in kindergarten seem not to diverge considerably in their school performances, especially when considering the overall migrant-native gap (there is still a small migrants' advantage at the bottom of the distribution of preschool achievement if we look at the net migrant-native gap).

These results are particularly intriguing. Why are high-achieving migrants in preschool outperforming natives with similar early achievement during primary education? A possible explanation is that high-achieving migrants in preschool are positively selected compared to high-achieving natives. This would be the case, for example, if Germany had a migration history in which immigrants, though negatively selected by SES, were positively selected by cognitive abilities. However, this interpretation is inconsistent with the existence of a migrants' preschool penalty in all competence domains (see Figure 4 and Figure 8), and it fails in explaining why migrants' advantages over the school years pertain high preschool achievers only. Still, it may be that immigrants in Germany are positively selected by cognitive abilities and that those higher cognitive abilities require some time to translate into higher learning outcomes. In this scenario, migrants performing as high as natives in kindergarten had a better cognitive endowment that would possibly cause migrants' better performances at later educational stages. However, this explanation is not consistent with our finding pointing towards a negligible difference in the basic cognitive abilities between migrants and natives (see Figure 4).

A third explanation is that migrants' have higher motivations which are hindered when they are more exposed to the family environment and tend to flourish when they spend time in the classroom. This explanation, however, does not elucidate why the migrants' advantage over the school years are found only among high achievers in preschool (see Figure 8) and why we do not find decreasing migrants' penalties in most competence domains (see Figure 4). Differences in the trajectories among specific groups of migrants would possibly help to find a plausible explanation, but this strategy is unfortunately not feasible in the German case due to data restrictions.

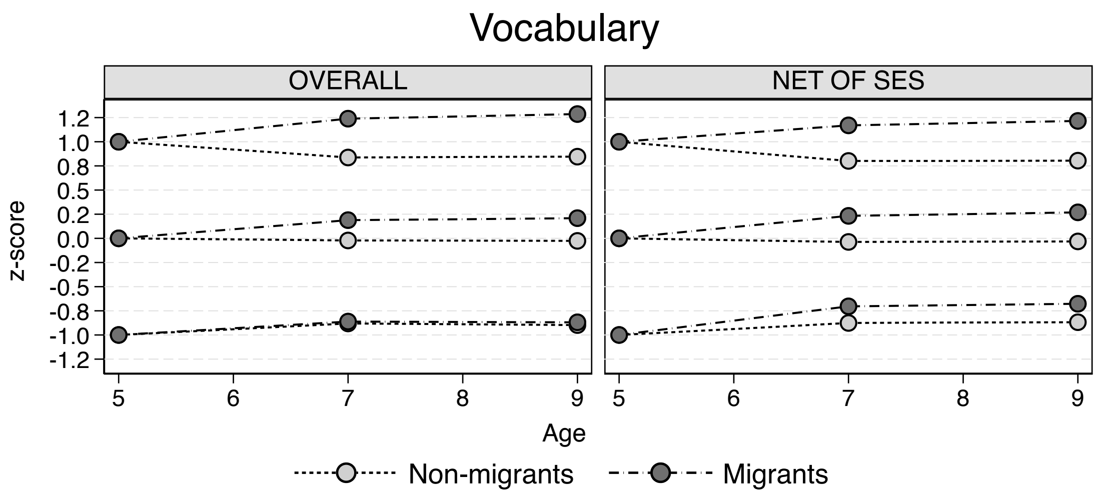


Figure 9 Simulated trajectories moving from preschool to the end of primary schooling of children starting with the same preschool achievement by migration background: overall (left panel) and net (right panel) migrant-native gaps.

Notes: Predictions based on life stage-specific regression models. Simulated trajectories reflect the z-scores we would observe in primary school if the additional SES effect beyond preschool depended on preschool z-scores (primary school z-scores predicted for children from different SES performing 1, 0 and –1 SD above/below the average preschool performance).

2.6 Conclusions

When do social and migration gaps in cognitive achievement emerge and how they evolve over the educational career? Are social and migration gaps in primary school fully explained by preschool inequalities or does family background play a role beyond preschool? By employing the most recent, representative, and rich competence data, the chapter examined social and migration inequalities in educational achievements in one of the countries with the more infamously stratifying education systems in the Western world: Germany.

We have documented profound differences in learning outcomes which align astoundingly along the social and migration background of children. However, in line with the idea that social and migration gaps in achievements are more environmentally than genetically determined, we found only a little difference in basic cognitive abilities. SES gaps arise as early as 7 months after birth, translate into remarkable achievement gaps even before children enter school, and remain – notwithstanding slight increases – fairly stable as children navigate through school. The evolution of gaps between the early cognitive achievements of migrants and natives is more dependent on the competence domain, however.

The migrants' penalty seems to decrease moving from the early years throughout schooling in some domains (science and orthography), while remaining stable (math) or even slightly increasing (reading) in others. The exceptionally pronounced migration gap in vocabulary, instead, seems to increase when entering Kindergarten and then to decrease dramatically over primary and secondary schooling. Interestingly, the vocabulary gap is consistent even among Russians, which perform much more similar to natives compared to Turkish. Indeed, Turkish lag dramatically behind both natives the other prominent group of migrants in Germany in all competence domains. What is more, while Russians seem to recover over the educational career in some domains, the performances of Turkish children seem to further depart from natives' performances along the educational cycle.

We have also shown that SES is an essential driver of the low performances of children with a migration background. However, while explaining a significant proportion of the migrants' penalty in most of the learning outcomes (between 50% and 100% depending on the domain), SES could not fully account for the lower performances in vocabulary tests. The latter result holds even in the case of children with Russian heritage, who tend to perform equally to natives when hailing from similar social backgrounds in all other competence domains. Generally, while the overall lower performances of Russians compared to natives seem to be more a story of social and economic deprivation, the astoundingly lower performances of Turkish cannot be solely traced back to economic and the social circumstances of their families of origin. A tentative estimation suggests that even if differences in the social and economic standing of Turkish and natives' families vanished, roughly 50% of the Turkish penalty in educational achievements would still be in place.

In-depth analyses of the competence domain in which social and migration inequalities are the highest – that is vocabulary – offered valuable insights on the role of early disparities for the formation of inequalities during the school career. Both SES and migration penalties in the mastery of German vocabulary are apparent in kindergarten, and then only slightly increase (SES) or decrease (migration) over primary schooling. However, we estimated that the 60-70% and the 100% of SES and migrants' penalties at the end of primary education respectively are attributable to inequalities that are well settled when children are not yet enrolled in the school system (in kindergarten).

On the one hand, it seems that around the 30–40% of SES inequalities in primary school emerge thanks to the additional role that social background plays beyond the preschool period, for example, by ensuring continuous support at home or taking advantage of non-formal primary school differentiations in terms of quality and student body composition. On the other hand, notwithstanding being associated with a significant penalty in the early years, children's migration background seems to act toward compensation of initial inequalities: when starting with the same performance in preschool, migrants outperform natives in primary school. Substantively, these results imply that without the additional role of SES and migration background beyond the early years of life, SES inequalities would increase more strongly over primary education than they increase nowadays, while migrants' penalties would remain stable rather than decreasing slightly.

Finally, we have also explored whether the role of SES and migration background beyond preschool concentrates among high or low preschool achievers. We found that high SES children performing poorly in preschool tend to recover over primary schooling compared to their low SES counterparts. Instead, high preschool achievers perform similarly in school irrespective of the social standing of their family of origin. While the additional SES effect beyond preschool concentrates among poorly performing children in preschool, the opposite holds for migration background. Migrants with good performance in preschool tend to outperform natives with similar initial performances, while poor preschool achievers – irrespective of whether migrants or natives – have similar trajectories afterwards.

Therefore, we can conclude that SES contributes to slightly increasing inequalities in the mastery of vocabulary over the early educational career by compensating for possible ‘bad starts’ in the early years. Conversely, migration background contributes to slightly decreasing inequalities by boosting the educational achievements of children already performing well in the early years. In this sense, migrants and low SES children performing poorly in kindergarten seem to have no shelters in the current situation and are possibly the primary targets for public policy interventions in contemporary Germany.

3 NETHERLANDS

The Evolution of Achievement Gaps from Early Childhood to Adolescence in the Netherlands

Thomas van Huizen
Utrecht University

3.1 Introduction

School segregation is comparatively high and rising in the Netherlands (Ladd et al. 2011; Boterman 2019; Inspectorate of Education 2018; Education Council 2018). An explanation for the high degree of social and ethnic segregation is the strong commitment to "freedom of education", safeguarded in the Dutch constitution for over a century. The education system is characterized by a high degree of parental choice, equal funding of public and private schools and a high degree of school autonomy. In the first part of the previous century there were no major concerns about social and ethnic segregation as religious denomination rather than social class or ethnicity determined school choice. However, increasing secularisation and the influx of migrants in the 1960's and 1970's resulted in higher levels of school segregation. Instead of reducing segregation per se (e.g. by limiting parental choice), the chosen policy objective was to provide high quality education for all children by alleviating the negative effects of segregation (Ritzen et al. 1997; Ladd & Fiske 2011). Since 1985 more public resources are allocated to schools with a higher share of disadvantaged children according to a weighted funded system. Moreover, substantial investments have been made in targeted preschool programs to alleviate skill disadvantages before school entry (Akgündüz & Heijnen 2018; Leseman et al. 2017). Hence, the Dutch education system aims to neutralize the key arguments against segregation by allocating additional resources to schools with high concentrations of disadvantaged children and by providing disadvantaged families access to relatively high-quality preschool programs.

The Dutch 'high segregation – high compensation' system is often considered successful in terms of efficiency and equity of achievement results: "The Dutch school system is one of the best in the OECD, as measured by the Programme of International Student Assessment (PISA) and the Survey of Adult Skills (PIAAC). It is also equitable, with a very low proportion of poor performers." (OECD 2016b: p.11). This result is especially striking because spending on education is not exceptionally high, indicating that the Dutch education system is rather efficient.[1] Recent international evidence based on PIRLS shows that in the Netherlands a comparatively small share of the variation in Grade 4 reading scores can be explained by school differences (UNICEF 2018). This tentatively suggests that the high degree of school segregation is effectively counteracted by the weighted funded system. Moreover, while SES gaps in Grade 4 reading scores in the Netherlands are higher than those in most Scandinavian countries, they are lower than many continental European countries (e.g. Germany, France, Belgium) (Rözer & van de Werfhorst 2017: p.55). However, several policy concerns remain: segregation may have negative

[1] Spending primary education is low (measures as a percentage of GDP) to average (measures as spending per student). Spending on secondary education is above the OECD average though (OECD 2018b).

consequences beyond achievement scores, e.g. in terms of social. Moreover, students are tracked in secondary school from around the age of 12. Although the jury is still out on the effects of tracking, limited permeability between educational tracks is an important policy issue (Inspectorate of Education 2018; Education Council 2018).

This chapter examines the evolution of achievement inequalities from early childhood to adolescence in the Netherlands. The central aim of this study is to complement the cross-sectional snapshots of achievement gaps by providing an answer to both research questions discussed in Chapter 1 of this report. First, I test when SES and migration-related gaps in achievement arise and how these gaps evolve during (early) childhood and adolescence. Are gaps already sizeable before children enter kindergarten and school? Do gaps narrow or widen as children move through primary and secondary school? Second, I examine the extent to which social and migration-related achievement gaps in primary school can be attributed to preschool achievement differences. Is there a substantial additional role of SES and migration background in the school years or are inequalities already settled in the years before school entry? Are low SES children with the same initial achievement as their high SES peers being left behind in primary school? Is there significant upward or downward mobility in achievement during the school years and does this vary by SES and migration background?

The results presented in this Chapter are based on two longitudinal datasets: Pre-COOL and COOL. Both datasets contain information on family SES and migration background. SES is measured by the level of education of the parents. Unfortunately, limited information about household income available. Moreover, migration inequalities are examined by estimating migrant-native gaps. In addition to overall migrant-native gaps, I examine whether specific migrant groups lag more behind than others, focusing on children with a Turkish or Moroccan background (two ISOTIS target groups).

COOL and Pre-COOL include data on a battery of achievement tests. The analysis focuses on inequalities in language/literacy and math/numeracy achievement as these two domains have been tested consistently in (Pre-)COOL. Moreover, from a life course (human capital) perspective these skills are highly relevant as language and math skills measured in childhood/adolescence significantly predict adult earnings and employment prospects (Chetty et al. 2014; Lin et al. 2016). Importantly, by combining the two related datasets, the evolution of gaps can be examined over an extended observation window (from age 2 to age 14).

Notwithstanding the strengths of the data, some serious limitations remain. First, as in Chapter 2, the analysis of the evolution of gaps relies on multiple cohorts. The data does only allow genuine longitudinal analysis from age 2 to 6 and from age 5 to 11 (and from 8 to 14). In line with the analysis presented in Chapter 2, a weighing strategy is employed to link different cohorts. Moreover, I discuss results that indicate that cohort effects in this time period are probably negligible. A second limitation is that, whereas COOL is a representative sample, Pre-COOL is not designed to be representative. Yet, by exploiting the overlap between the two datasets (i.e. kindergarten achievement results), weights can be assigned to correct for non-representativeness of the Pre-COOL sample. When comparing COOL and Pre-COOL results, the estimates appear to be very consistent with each other. Finally, the major limitation is probably panel attrition. Although longitudinal weights are used to limit concerns associated with non-random attrition, the sample size for longitudinal analysis is relatively small as a consequence of panel attrition. This implies that the study of research question 2, explaining school gaps by preschool differences, is limited to SES and overall migrant inequalities and does not distinguish between different migration groups.

The remaining of the chapter is structured as follows. The next section provides an overview of the Dutch institutional setting. In the third section, the data, measures and methodology to address the main research questions are discussed. Subsequently, the fourth section presents the empirical results. The final section concludes and discusses several policy implications.

3.2 Institutional context

3.2.1 Family policies

With respect to family policies, the Netherlands stands out from other continental European countries. For instance, in most continental European countries public spending for families with children is substantially above the OECD average, whereas it is below average in the Netherlands (Thévenon 2011). Maternity and parental leave policies are not generous for European standards: post-natal maternity leave is three months (100% replacement rate) and the duration of parental leave is 6 months. However, parental leave entitlement is on a part-time basis and it is often unpaid (the level of payment, if any, is regulated in the collective labour market agreements). Furthermore, full-time childcare attendance is rare in the Netherlands (see 3.2.2). Hence, the Netherlands shares some features with the UK, combining short leave with part-time childcare services.

A unique feature is that The Netherlands is “the first part-time economy in the world” (Visser 2002: p.23) and this is reflected in the rather unique solution to reconcile work and family life (Plantenga 2002). Female (maternal) employment is relatively high, but around three out of four working women – not just mothers with young children – are employed on a part-time basis. Given the short leave duration, it is common for mothers in the Netherlands to return to employment within the first half year after childbirth, but only for a limited number of hours. In the years before the child enters kindergarten at age four, the typical solution is part-time employment of mothers (reduction of working hours of fathers is also not uncommon), around two days of formal day care and one day of informal care by grandparents.

3.2.2 Early Childhood Education and Care

3.2.2.1 Participation in ECEC

Given the ‘Dutch part-time solution’, most Dutch children spend some time in ECEC before entering kindergarten. ECEC participation rates in centre-based care and regulated family day care are high for children below age three: 56% versus the EU average of 32%. Only Denmark has a higher ECEC participation rate for this age group (OECD 2016a).[2] As in all other EU countries, participation rates are higher for children aged three to five (92% in the Netherlands). Because kindergartens are free and universally available (see Section 3.2.2), participation jumps from below 81% for children aged three to over 96 for children aged 4 (at age 5 participation is

[2] As in most EU countries, ECEC participation rises sharply with age: whereas 23% of children aged 0 participates in (centre/family) day care, this increases to 56% for children aged 3 (Statistics Netherlands 2016).

mandatory).

In the Dutch case the ECEC attendance rates do not provide an accurate indication of the total exposure (intensity x duration) to ECEC as children are typically enrolled in childcare for a limited number of days per week. The modal use of centre day care is two days a week and the average ECEC hours is around 16 hours for the 0–2 group (which is comparable to the UK). In fact, only a small percentage of Dutch children spend 30 hours or more in ECEC: 6% of children aged 0–2 and 14% of children aged 3–4 (OECD 2016a). This implies that in terms of full-time equivalents, the ECEC participation rate of 0/2-year-olds is below the EU average (32.8 versus 36.7). Hence, ECEC participation rates for children below age three are among the highest in EU but below the EU average in the full-time equivalents.

As in many European countries, the Dutch ECEC system is organized as a split system, with different arrangements for children until age four and for children aged four until the start of primary school. For children below the age of four, various child care arrangements exist; some services are aimed at supporting work and family life and encouraging female labour market participation (day care centres), whereas others have typically a stronger focus on education (playgroups). An important feature of the Dutch educational system is that it has specific policies to reduce educational disadvantages at an early age.

3.2.2.2 The Dutch ECEC system: Alternatives, eligibility and funding

Day care centres (*kinderdagverblijven*) provide childcare services for zero- to four-year-olds and are the most commonly used facility for children in this age range. Centres provide full-day care for five days a week and primarily offer services to dual-earner families and are for that reason under the responsibility of the Ministry of Social Affairs and Employment. Since the introduction of the 2005 Child Care Act (*Wet Kinderopvang*), the day care market has been privatized and both commercial (for profit) and non-profit organizations operate on the market (see Noailly & Visser (2009) and Akgündüz & Plantenga (2014a) for a more extensive discussion of the reform). The Dutch day care system stands out as one of the few European countries without public provision of day care. The financing system is demand-driven: parents can select their preferred centre, conditional on availability of slots, and receive an income-dependent subsidy through the tax system. Although prices are not strictly capped, there is a soft price cap as there is a maximum hourly fee that is subsidized (7.18 euros in 2017). Day care prices are therefore generally not substantially higher than this maximum price. The subsidy system has been reformed several times since 2005; initially subsidies increased, resulting in drop in average parental costs from 37% in 2005 to 18% in 2007 and, consequently, an expansion of the day care sector: public spending on childcare subsidies tripled over the period 2004–2009 to 3 billion euro (Bettendorf et al. 2015). The subsidies were reduced somewhat in 2009 and more substantially in 2012. During the more recent years, subsidies have become more generous again. In the recent period, parents paid around one third of the gross price: net prices for the lowest income groups are 6%, whereas higher income groups pay up to two thirds of the gross price.[3] Subsidies for the second child in day care are substantially higher. It should be noted that actual day care costs for parents are rather uncertain ex ante, as the subsidies are based on realized (ex post) annual household income. Importantly, families are eligible for child care subsidies when both parents in the

[3] These figures refer to 2017. Given that subsidies are set for a maximum hourly price, actual parental costs may be higher.

household are employed, in education or actively looking for work; single-parent families are also eligible when the parent is employed, in education or actively looking for work. However, breadwinner families are not eligible for day care subsidies.

As an alternative to day care centres, working parents can also family day care services (*gastouderopvang*). Although less popular than day care centres, family day care services are a nontrivial part of the Dutch ECEC landscape, with around 7% of children aged 0–3 participating in these arrangements. Family day care is also regulated and subsidized and can be considered as formal, non-centre based ECEC. In case these facilities are registered in the national child care register, the same subsidy conditions apply as those for day care centres. Fees for in-home care services are generally lower than fees for day care centres.

In addition to day care centres and in-home services that provide full-day care for zero-to four-year-olds, playgroups (*peuterspeelzalen*) offer a part-day and more formal type of ECEC for children between age two and half and age four (before entering kindergarten). For children in the relevant age range, playgroups represent an important alternative to centre/family day care: 28% participates in playgroups, compared to the 43% in day care (and 9% in both types) (Statistics Netherlands 2015). Playgroups are run by municipal welfare organizations, day care centres and primary schools. Playgroups only provide half-day programs (2.5 hours per day), typically two days per week for around 40 weeks per year. Given the limited operating hours, the aim of playgroups is not to facilitate parental employment but rather to prepare children for kindergarten and school. Instead of subsidies through the tax system, playgroups are subsidized by municipalities (supply subsidies) and parents pay an income-dependent fee determined by the municipality.[4] Since the implementation of the Day Care and Playgroups Harmonization Act in January 2018, there are no formal differences between day care and playgroup organizations and the latter as a type of organization no longer exist. However, the type of service (i.e. an ECEC program of two half-days) are still offered by day care organizations, including the organizations formerly registered as playgroup. During the years before the 2018 reform there was a significant decline in the number of preschools.[5] In fact, most preschools formally changed from preschools to day care centres, providing de facto the same service but allowing dual-earner to receive subsidies for preschool services.

Furthermore, next to day care centres, in-home care services and playgroups that are in principle universally accessible, an important feature of the Dutch ECEC system is that it includes preschools targeted towards disadvantaged children. Since the 1970's, the Netherlands introduced policies that aim to actively reduce educational disadvantages. In the current educational setting, the Dutch Educational Disadvantage Policy (*Onderwijsachterstandenbeleid*) includes pre- and early school programs (*Voor- en Vroegschoolse Educatie*; *VVE*): preschool programs for children aged two-and-a-half to four, and early school programs for four- and five-year-olds. The latter are part of kindergarten and are discussed below. The central goal of these targeted preschools is to reduce early educational and developmental disparities. The preschool programs consist of 10–12 hours (four half-days) centre-based ECEC per week for around 40 weeks.[6] Programs are provided by day care centres or (before the 2018 reform) playgroup

---

[4] For instance, in 2018 in Utrecht the annual parental fees varied between 45 and 790 euro, depending on the level of household income.

[5] The number of preschool organizations dropped by almost 40% between the end of 2013 and early 2017 (Buitenhek, 2017).

[6] Municipalities are legally responsible for the provision of preschool programs of (at least) 10 hours per week for

organizations. The national budget is allocated between municipalities using the primary school weights system. Essentially, municipalities receive a larger part of the budget when they have more primary schools with a high proportion of children with low educated parents. The total preschool budget has increased considerably during the last decade, from 200 million euros to 260 million euros in 2011. Furthermore, substantial additional funding was allocated to the 37 largest municipalities (70 million in 2012 and 95 million in 2013) (Akgündüz & Heijnen 2018). Municipalities use this funding to provide subsidies to centres offering preschool programs and have a certain degree of autonomy concerning their preschool policies. In most municipalities eligible children can participate in preschool for free or for a small parental fee. Municipalities also differ in their targeting (eligibility) criteria, although the educational background of parents generally plays a major role.

3.2.2.3 Inequality issues

The actual use of ECEC is strongly related to family SES. There is a significant difference in the use of ECEC services for 0–2 aged children between families with higher and lower educated mothers: 46% (no tertiary education) versus 67% (with tertiary education) (OECD 2016a). Concerning 2/3-year-olds, available data indicates that ECEC participation increases substantially by household income: 92% of children in the highest household income quintile participate in ECEC versus 62% in the lowest income quintile, i.e. almost 4 out of 10 children in the lowest income group do not participate in ECEC before entering kindergarten (Statistics Netherlands 2015).

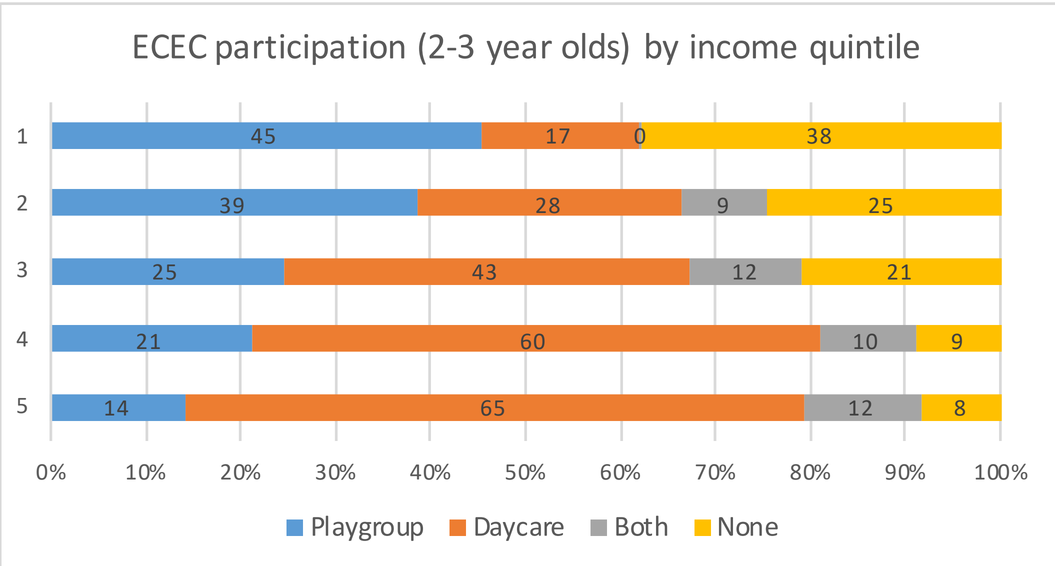


Figure 1 ECEC participation (2-3 years olds) by income quintile.

Source: Statistics Netherlands (2015).

---

disadvantaged children.

Families with higher income are not only more likely to use ECEC services, they also opt for different types of ECEC services. As shown in Figure 1, the propensity to use playgroups declines with household income, whereas the propensity to use day care increases with household income. This can be expected as the lowest income households are often not eligible for subsidies; these families are generally breadwinner families or families with no employed parent).

Empirical findings based on Pre-COOL show that equality of ECEC quality is to a large extent achieved in the Netherlands. First, preschools offer higher quality than day care centres. Second, ECEC providers with a higher share of disadvantaged children provide higher structural and process quality: "These results indicate that the targeted preschool policy in the Netherlands succeeds in providing higher quality ECEC to those who need it most" (Leseman et al. 2017: p.182). Results from multivariate analyses that take into account selection into ECEC generally find no major gaps in ECEC quality accessible to different groups (Akgündüz & Plantenga 2014b). Overall, these results are somewhat more mixed. On the one hand, higher SES families seem to use ECEC services with higher emotional support. On the other hand, native families use services of lower emotional support compared to migrant families.

While there is no causal evidence available on the impact of participation of targeted preschools, some results tentatively suggest that targeted preschools effectively reduce disadvantages. First, results based on Pre-COOL indicate that disadvantaged children catch up significantly compared to non-disadvantaged children in vocabulary and selective attention (executive function). However, it is unclear whether this can be attributed to participation in preschools. As Leseman and colleagues (2017: p.186) note: "The design of the pre-COOL study does not allow for strong conclusions about the effectiveness of participating in centre-based ECEC for developmental and educational outcomes. The catching-up effects that were found cannot be attributed unambiguously to participation in ECEC because no meaningful comparison could be made with equivalent children without any ECEC experience." Second, Akgündüz & Heijnen (2018) follow a quasi-experimental approach (difference-in-differences), exploiting a reform that increased funding for targeted preschools substantially for the 37 largest Dutch municipalities. The results show that grade repetition in kindergarten – an indicator for school readiness – declined by .8–1.8 percentage points for the targeted population. This effect is substantial given a grade retention rate of around 10% for the specific group.

3.2.3 Kindergarten and primary school

3.2.3.1 Central features

Dutch primary schools provide education from 4- to 12-year old children and consist of 8 'groups' (years), where group 1 and 2 provide a kindergarten program and are often mixed in terms of age. Formal schooling in primary schools starts in group 3 (around age 6; Grade 1). While in many other European countries kindergartens operate to a large extent independently from primary schools, in the Netherlands kindergartens are completely integrated in the primary school system. This implies that in the Dutch context kindergartens are formally part of the primary school system rather than the ECEC system. Hence, issues related to school choice, funding, (structural) quality features etcetera, apply equally to kindergartens as to the other primary school grades.

The Dutch system can be described as a highly decentralized system, balanced by accountability (OECD 2016b; Ministry of Education, Culture and Science 2016). While both public and private schools exist (the former caters to around one third of primary school-aged children), almost all primary schools are publicly funded.[7] Private schools are generally based a particular religion (e.g. Protestant, Islamic) or educational model (e.g. Montessori). Every citizen has the constitutional right to start a private school and receive government subsidies. In fact, equal public funding rules apply for private and public schools.

Schools have a large degree of autonomy on how they allocate funding: a so-called 'lumpsum funding system' is in place. In addition to the number of pupils, funding rules take into account the composition of staff (funding increases with the age of teachers), the educational level of parents (schools with a high concentration of disadvantaged children receive more funding) and the school neighbourhood (schools in specific deprived areas, '*impulsgebieden'*, receive additional funding). Schools have a high degree of autonomy in how they allocate the public funding. For instance, the additional funding for a high share of disadvantaged children is not earmarked. Furthermore, schools also have a large degree of autonomy over the teaching content and method. There is, for example, no national curriculum: "'Freedom to organise teaching systems' means that both public and private schools are free to determine what is taught at schools and how this is taught, within legal boundaries." (Ministry of Education, Culture and Science 2016: p. 1)

The large degree of autonomy is balanced by national quality standards and examinations and a strong inspectorate of education. These standards include teacher qualifications, subjects to be taught, attainment levels but also the content of national examinations. The inspectorate of education plays an important role in monitoring schools. The inspectorate evaluates school quality and classifies schools performing below standard as (very) weak schools. Weak schools will be monitored closely; when they do not adhere to the standards in the following assessment funding will decline or schools will be closed.

3.2.3.2 Kindergarten

When children turn age four, they are entitled to enrol in kindergarten and almost all children do so. Participation is mandatory from age five. Given that they are part of the primary school system, kindergartens are free and universally accessible. Programs are for 25 hours per week, five days a week during school weeks. Kindergarten teachers are essentially primary school teacher in terms of qualification requirements (Bachelor degree) and wages. To facilitate parental employment children can use subsidized out-of-school care before or after the operating hours of kindergarten programs. There is no standard duration of kindergarten as children generally enrol during the school year (i.e. when they turn four). Formal schooling starts in 'group' 3 (Grade 1 in according to international definitions) and the transition to this grade depends on the teacher's assessment of the child's school readiness. This implies that some children may spend less than 18 months in kindergarten, whereas others may spend over three years in kindergarten. Grade retention in the second year of kindergarten ('*kleuterbouwverlenging*') occurs when the child is not considered 'school ready' and is relatively high (around 7%).

As mentioned above, targeted policies include early school programs for disadvantaged children. The total budget for these programs offered in kindergarten is smaller than the preschool budget (50 versus 200 million in 2010). However, in addition to these programs, the funding of

[7] Less than .5% of students are in private, not publicly funded primary schools (OCW 2016)

primary schools depends on the school's student composition. As kindergartens are part of primary schools, schools with a higher share of disadvantaged children receive more funding and can hire for instance more (specialized) kindergarten staff.

3.2.3.3 End of primary school

When children leave primary school around age 12, children enrol in a one of the three main secondary school tracks (see Section 3.2.4). Track choice is determined by the school track recommendation and an independent, standardized test. Scores of these tests correspond to a specific track recommendation. A recent reform has increased the relevance of the school recommendation. Before 2014–2015, teachers formulated their recommendation when the results of the standardized tests were available and in the large majority of cases the recommendation was consistent with the test results. Both the school recommendation and test results played a role in the actual track placement. Since 2014–2015, the school recommendation has become dominant for track placement. Teachers formulate their recommendation before test results are available; this recommendation should be reconsidered if test results indicate a higher track.

3.2.3.4 Inequality issues

The Netherlands has a universal kindergartens integrated in the primary school system. Evidence indicates beneficial effects of Dutch kindergartens, especially for disadvantaged children. Exploiting exogenous variation in kindergarten enrolment opportunities, Leuven et al. (2010) report that increasing the enrolment opportunities significantly improves achievement scores. Consistent with the literature on ECEC effects before kindergarten (van Huizen & Plantenga 2018), disadvantaged children benefit from increased kindergarten participation while non-disadvantaged children do not benefit. This finding implies that Dutch kindergartens substantially reduce achievement gaps: increasing the kindergarten enrolment opportunities by one month reduces the gap between disadvantaged and non-disadvantaged children by almost 10% (given a gap of .6–.7 SD).

A second important feature of the Dutch system is the rather high school segregation, which can probably be attributed to the high degree of parental choice and the high degree of school autonomy. In this policy context, higher educated parents appear to navigate carefully through the educational landscape. A comparison with the US indicates that segregation in the bigger cities in the Netherlands is at least as high as in large US cities (Ladd et al. 2011). Evidence also shows that segregation is increasing over time (Inspectorate of Education 2018). High SES families tend to opt for schools with specific educational programs (e.g. Montessori), that often ask for high voluntary parental fees.[8] Moreover, the existence of specific religious schools (e.g. Islamic schools) may contribute to ethnic segregation (Inspectorate of Education 2018). Nevertheless, descriptive evidence does not suggest clear negative consequences of segregation in terms of achievement (inequality). For instance, migrant achievement gaps declined between 1994 and 2004; SES gaps remained fairly stable (Ladd & Fiske 2011). Moreover, according to the most recent wave of PIRLS, the Grade 4 achievement gaps between

---

[8] Although Dutch schools are free, schools may ask for voluntary parental contributions to finance extra-curricular activities (e.g. school trips, music lessons). When families do not pay these contributions, their children can be excluded from extra-curricular activities but not from core education activities.

low and high achievers (10th versus 90th percentile) are the smallest in the Netherlands (UNICEF 2018).

3.2.4 Secondary school

3.2.4.1 A tracked system

The Dutch secondary education system contains three different tracks: 1) vmbo: pre-vocational secondary education. This track takes four years to complete and gives access to MBO. The vmbo track is divided into four different levels or subtracks; 2) havo: senior general secondary education. This track takes five years to complete and gives access to HBO (higher vocational training, leading to a Bachelor's degree); 3) vwo: pre-university education. This track takes six years to complete and gives access to WO (universities, leading to a Master's degree).

Some schools offer a combination of tracks (sometimes at different locations), other schools offer only one type ('categorial schools'), for instance only vmbo or vwo. The former type of schools generally offers more comprehensive classes in the 1st grade or first two grades of secondary schools, the so-called 'bridge classes'. These bridge classes consist of a combination of tracks. For instance, a common system is to offer a one-year vmbo/havo class and a two year havo/vwo class. The lowest achieving students in the first year vmbo/havo class will be tracked in a vmbo track, while the better performing students will move to a havo class or join the second year havo/vwo bridge class. Given that bridge classes are very common in the Netherlands, mobility in the first and second year is substantial: for most students the final track is clear in year three of secondary school. Around 25% moved to another track (Inspectorate of Education 2015).

3.2.4.2 Inequality issues

In general, low SES children and children with a migration background are more likely to enrol in a lower secondary school track. However, this can be expected as track placement in the Netherlands is to a large extent achievement based and these groups on average obtain lower achievement scores. However, gaps remain generally significant when conditioning on end of primary school standardized test scores (e.g. Inspectorate of Education 2018). Theoretically, the residual gap could be explained by measurement error in achievement, differences in non-cognitive skills, parental choice or discrimination. Hence, it is not obvious that this is evidence of inequality of opportunities. Studies relying on international data (e.g. PISA) produce mixed results on the effects of tracking (Brunello & Checchi 2007).

An interesting feature of the Dutch tracking system is the existence of bridge classes, as they facilitate mobility between tracks in the first year(s) of secondary school. Quasi-experimental evidence shows that a lack of opportunities to enrol in a bridge class (due to lower local supply of comprehensive schools) negatively affects the probability to complete higher education. This effect is larger for higher SES pupils (van Elk et al. 2011). Given this result it seems worrying that the use of bridge classes has declined in the past decade (Education Council 2018).

3.3 Data and methodology

3.3.1 (Pre-)COOL: Design and sample

The findings presented in this Chapter are based on the Dutch Pre-COOL and COOL data. Both datasets include a range of competence tests as well as information on family background. As in Chapter 2 (Germany) and 5 (UK), the analysis takes a long-term perspective (age 2–14) in the analysis of the extent and evolution of achievement gaps (RQ1). The analysis focuses on the early years (age 2–6) using Pre-COOL (2010/11–2014/15) and relies on data from the most recent wave of COOL (2013/14) to estimate the evolution of skill gaps until age 14. Hence, this approach combines a longitudinal cohort design (age 2–6, Pre-COOL data) with a pseudo-cohort design (age 5–14, COOL data). Our results for RQ1 are based on the full unbalanced sample.

The analysis explaining end-of-primary school achievement gaps by preschools achievement differences (RQ2) requires longitudinal data. These analyses therefore rely on the balanced sample. COOL is suited for this purpose as it follows children from kindergarten to the end of primary school.[9]

3.3.1.1 COOL$^{5–18}$

COOL$^{5–18}$, Cohort Study Educational Careers of 5/18-year-olds (*CohortOnderzoek OnderwijsLoopbanen*; COOL) has been collected triennially from school year 2007–2008 to 2013–2014.[10] The sampling design of COOL is school-based, implying that children are sampled from grades rather than cohorts (similar to NEPS, see Chapter 2). This means that the results are a better reflection of stage-specific inequalities than age-specific inequalities.

The data collection was executed by a larger consortium: ITS and the Kohnstamm Instituut were responsible for the part concerning kindergarten/primary schools, whereas Cito and GION were responsible for the part concerning secondary schools and MBO (age 14–15/17–18). COOL is based on a multi-cohort sequence design and includes in total 6 cohorts: 4 cohorts in COOL1, plus one additional kindergarten cohort in COOL2 and one in COOL3 (see Table 1).

The kindergarten and primary school data consist of a representative school sample (400 schools/almost 28,000 children in COOL1) and an additional sample of schools with a high share of children from disadvantaged backgrounds (150 schools in COOL1). Given the general aim of this report I rely on the representative sample: this sample cover around 6% of schools in the Netherlands. The collection of the representative sample took into account the national distribution of schools (using national administrative records) in terms of social-ethnic composition (6 categories), denomination (4 categories: public, Protestant, Catholic, other), province (12), degree of urbanization (5 categories). With respect to these four dimensions, the actual distribution of schools in the representative COOL sample does not significantly differ from the national distribution of schools.

Although the set-up of COOL is longitudinal, which allows children to be followed for several years, a substantial part of the sample drops out of COOL between consecutive waves. Around 44% of COOL2 Grade 3 children participated in COOL1 (kindergarten) and almost 60% of COOL3 Grade 6 children participated in COOL2 (Grade 3). There are three major reasons for panel attrition: schools do not participate in COOL anymore[11]; grade retention; or children move to another school.

Finally, in addition to kindergarten and primary school data, COOL contains information

[9] This is not (yet) possible with Pre-COOL data.

[10] See Driessen et al. (2009) for more information on the sampling procedures of COOL.

[11] For instance, around 40% of schools that participated in COOL1 did not participate in COOL2.

on achievement of children in the third year of secondary schooling (age 14–15). However, a sample-population comparison of the distributions in terms of school track, region and degree of urbanisation indicates that the secondary school data is not fully representative of the population (Zijsling et al. 2017). For instance, children in the vocational (vmbo) track are underrepresented (and those in the havo and vwo track overrepresented).[12] As no weights are provided it is not straightforward to correct this. The results on achievement gaps in secondary schooling should therefore be interpreted cautiously. COOL also includes achievement data for individuals aged 17–18. However, the timing of data collection beyond age 14 is track-specific and therefore no overall comparisons can be made for the oldest COOL group.[13] In this report I therefore rely on the age 5–14 data.

Table 1 Overview of the COOL waves, cohorts and samples.

			COOL1 07/08	COOL2 10/11	COOL3 13/14
Sample: number of schools					
Kindergarten/ primary schools			400	406	340
Secondary schools			81	151	107
Sample: number of children					
Stage	Age				
Kindergarten	5	Cohort	C1-K07/08	C2-K10/11	C3-K13/14
(year 2)		N	10069	9261	7279
Grade 3	8	Cohort	C1- K04/05	C2- K07/08	C3- K10/11
		N	9288	10109	7449
Grade 6	11	Cohort	C1-K010/2	C2- K4/5	C3-K07/08
		N	8545	9444 (12538)	7907
Secondary school year 3	14	Cohort	C1-K98/99	C2-K1/2	C3-K4/5
		N	8884	21384	16297

Notes: For RQ1 we rely mainly on COOL3. The analysis of RQ2 is based on the kindergarten 2007/08 cohort (cells shaded grey). Sample size refers to the representative sample; numbers indicate gross sample size.

[12] In COOL3, 38,6% of the sample follows a vmbo track, whereas this is almost 54% in the population.

[13] Age 18 for those in the vwo track (6th and final year of vwo), age 17 for those in the havo track (5th and final year of havo), and age 18 for those in mbo.

3.3.1.2 Pre-COOL

Pre-COOL is a cohort study that includes rich information on child development in the early years. The Pre-COOL sample includes in total over 3000 children who were age 2 in 2010 and consists of two subsamples: a center-based cohort and a home-based cohort.[14] Child development is assessed annually from age 2; average age was around 2.6 years at the first assessment in 2010–2011. For the analysis I use all five currently available data waves (2010–2011/2014–2015). Given our research design it is important to note that the fourth Pre-COOL wave (2013–14) overlaps with COOL3.

The existing COOL data collection infrastructure was taken into account when recruiting the Pre-COOL participants.[15] COOL schools that were willing to participate in Pre-COOL ('Pre-COOL schools') played a central role in the recruitment of both subsamples. First, the recruitment of the centre-based cohort occurred mainly via Pre-COOL schools. Schools were asked which ECEC providers (day care centres, playgroups) most children participated in before enrolling in the specific primary school. Pre-COOL centres are therefore generally located near Pre-COOL schools. Second, the home-based sample is based on a random sample draw by Statistics Netherlands in the postcode areas of Pre-COOL schools. The home-based cohort includes both children participating in ECEC and children not participating in ECEC. Additional recruitment effort was undertaken to increase the participation of children with migrant parents, including parents with a Moroccan and Turkish background.[16] Although Pre-COOL subjects are generally well-spread geographically (both rural and urban areas; all provinces are covered), Pre-COOL is not designed as a representative sample. However, as demonstrated below, after assigning weights based on the representative COOL sample, results from Pre-COOL are highly consistent with results from COOL.

3.3.2 Achievement measures

The central aim of this study is to examine how achievement gaps evolve over an extended observation window. For our main analysis I therefore rely on language and math achievement test data, as these are to a large extent comparable in the Pre-COOL (age 2–6) and COOL (age 5–14).

3.3.2.1 Pre-COOL

Pre-COOL contains two types of assessment data. First, a test battery has been developed for Pre-COOL to measure child development in the early years. These tasks were administered by trained research assistants. The rich data on the early years is one of the strong features of the Dutch data. Several domains have been assessed consistently from age 2 to 5 using the same task, increasing the difficulty with the age of the child.[17]

Second, for children aged 4 to 6 (kindergarten until 1st grade), Pre-COOL includes data

---

[14] In addition to the age 2 cohort, Pre-COOL includes an age 4 cohort. However, the latter sample is much smaller (around 700-800) and contains relatively limited child development measures.

[15] See Mulder et al. (2014) and Slot et al. (2015) for further details on the sampling procedures.

[16] For instance, families were visited at home by Berber and Arabic-speaking research assistants.

[17] For instance, receptive vocabulary has been tested every year from age 2 to 5 using a short version of PPVT.

from the Cito child achievement monitoring system (Cito-Leerlingvolgsysteem; Cito-LVS), administered by primary schools. In this monitoring system children can be tested twice a year (mid- and end-year test results are available). The Cito Language for Toddlers (Taal voor Kleuters) and Math for Toddlers (Rekenen voor Kleuters), the two tests administered in kindergarten, are both rather comprehensive tests. Cito Language for Toddlers measures language development (conceptual awareness, including receptive vocabulary) and emergent literacy (metalinguistic awareness).[18] Cito Math for Toddlers captures number knowledge, measurement and geometry. Both tests consist of two parts that are administered on two separate days and each part requires 20 to 30 minutes to complete. Several studies indicate good test reliability.[19]

Since the Cito monitoring system is used often by Dutch primary schools this data could often be supplied with limited additional effort from schools. However, not all Pre-COOL schools used this monitoring system and some schools administered the test only once or twice in kindergarten (there are four test moments in kindergarten). Test data is most complete for kindergarten mid-year 2; this is relevant as COOL also contains the mid-year Cito test results (see below).

For the analysis I use both Pre-COOL test battery data and Cito test data (see Table 2). For the description of the evolution of gaps over an extended observation window (RQ1) I focus on tasks that have been consistently administered from age 2–5 in the Pre-COOL test battery and that are related to competence assessments administered in COOL. For language, this concerns four tasks: PPVT (receptive vocabulary), phoneme task; nonword repetition task, grammar task. However, due to a rather low percentage of task completion of the latter two tasks, our main results are based on the first two tasks. Additional analyses include the nonword repetition task and grammar task. For math, I use the Cito Math for Toddlers tests (age 3–5).

3.3.2.2 COOL$^{5–18}$

Almost all tests used in COOL are from the Cito monitoring system.[20] Given that most schools use these tests and the associated software to process the test data, providing the test data to COOL implied marginal additional effort from these schools. If schools do not use the Cito monitoring system, the researchers provide the test. Given that using these tests is an integral part of most schools’ policies and therefore common practice, non-response is low: depending on the wave and stage, 94–98% of children participate in at least one test. Non-response in test scores can be due to absence on the test day (due to illness) or because the child moved to another school between the date schools supplied enrolment information and the test date.

Table 2 provides an overview of the COOL test data and shows where COOL and Pre-COOL overlap. In the kindergarten and primary school phase, test data is collected for three stages: second year of kindergarten (K–2; age 5), Grade 3 (age 8) and Grade 6 (age 11). For the analysis of gaps in the language domain I use the following language test data:

---

[18] In year one of kindergarten the emphasis is on conceptual awareness.

[19] See Lansink and Hemker (2012) for more details on the Language for Toddlers test and Janssen et al. (2005) and Koerhuis and Keuning (2011) for more information on the Math for Toddlers test.

[20] An exception is the NSCCT (Non-School Cognitive Capacity Test; Niet-Schoolse Cognitieve Capaciteiten Test), a test that is similar to IQ test. This data is only available for children in Grade 3 and 9 (year 3 of primary school).

- Kindergarten (second year): Language for Toddlers (as in Pre-COOL, see above).
- Grade 3 and Grade 6: Vocabulary; Reading comprehension. In the main models I rely these three tests to generate a composite language measure.
- Grade 6: Cito End test, language component. This data is only available when schools administer this test as a regular part of school activities. In the relevant time period, around 85% of children made the Cito End test. Because some schools use alternative tests and the decision to do so is probably non-random, this sample may be not be fully representative (although our weighting strategy may to some extent account for this). I therefore consider the composite measure discussed above as the main measure for Grade 6 language achievement.
- Secondary school (third year): COOL includes a Math test (some use Math test developed by Cito); different versions with overlapping items depending on track: IRT equivalent scale.

For the description of gaps in the math/numeracy domain I use Cito Math for Toddlers (age 5; as in Pre-COOL, see above) and Cito Math tests for Grade 3 and Grade 6. Similar to the analysis of the gaps in the language domain, I present results using the math component of the Cito End test.

Table 2 Assessment data: extent and evolution of language and math gaps (RQ1).

Domain	Age	Measure/test	Data	
			Pre-COOL	COOL3
Language	2-6	Composite measure (Pre-COOL test battery): vocabulary, phoneme awareness, grammar *, verbal short-term memory (nonword repetition) *	X	
	4	Cito test: Language for Toddlers	X	
	5	Cito test: Language for Toddlers	X	X
	6	Composite measure (Cito): vocabulary, reading speed/accuracy	X	
	8/11	Composite measure (Cito): vocabulary, reading comprehension		X
	11	Cito End test – language component		X
	14	Composite measure (Cito): reading comprehension, grammar/orthography		X
Math	3	Cito Math for Toddlers (short version)	X	
	4	Cito Math for Toddlers	X	
	5	Cito Math for Toddlers	X	X
	8/11	Cito Math		X
	11	Cito End test – math component		X
	14	Math		X

Notes: * Given the rather low task completion rate, we exclude these from the main to remain. However, including these additional tests does not substantially change the results.

As most of the tests are part of the child achievement monitoring system, it is important to note that the test results play a crucial role in the educational careers of children. The school recommendation and end of primary school tests (in particular the Cito End test) determine whether someone has access to specific secondary school tracks. When formulating the school recommendation, teachers take into account these test results.

3.3.3 Socio-economic status and migration background

Socio-economic status (SES) of the child's family is measured using information on parental education.[21] This information is gathered via two sources in COOL: a survey directly administered to parents and school registry data. The survey information provides more detailed information on parental education. However, as the survey response rate is around 70%, relying only on the parental survey data implies a substantial drop in the number of observations. Moreover, the analytical sample may not be representative given that non-response may be non-random. A strong feature of COOL is that in addition to data from the parental survey, schools supply information on parental background. Schools often have this information as funding depends on the educational level of children enrolled in school (see Section 3.2.3). In fact, schools have a financial incentive to register this information, especially when the parental background is low. School registry data is less detailed but more complete (over 95% of the sample). Based on the parental background information supplied by schools it appears that survey non-response is indeed non-random: non-response is relatively low among low educated parents and migrant parents. Using both sources of parental education, the SES level for over 98% of the children in the COOL sample can be determined.

As in COOL, Pre-COOL also used multiple sources to complete data on family background: parent questionnaires, centre records and school registry data (as in COOL). The response rate for the first parent questionnaire was rather low (83% for the home-based cohort and 45% for the centre-based cohort).[22] Questions on the educational level of the parent were therefore also included in questionnaires of subsequent waves. When combining the information from different sources, a SES level can be assigned to about 80% of the children in the sample.

Following Bradbury et al. (2015) and the other chapters of this report, SES is measured by the highest qualification obtained by the child's parents. I use the information from the parental survey and complement this with data supplied by schools. In case information for one of the parents is missing, I use data on the highest level of education of the other parent. As in the other chapters of this report, I distinguish between three SES categories (see Table 3): 1) High: the required years of education is 15 or higher. At least one parent holds a bachelor's degree or a higher degree; 2) Medium: the required years of education is between 11 and 14. The lion's share of this group completed vocational education (mbo or equivalent); 3) Low: parents who completed the vocational track of secondary school (or lower educational attainment), requiring maximum 10 years of education. The distribution is roughly 43% (high), 40% (medium) and 17% (low).

Data on migration background was also collected through multiple sources. In addition to

[21] In both Pre-COOL and COOL indicators for household income are incomplete and imprecise.

[22] This difference can be explained by the fact that families participating in the home-based cohort were visited at home. In contrast, the questionnaires were distributed to the families by the centre and families were asked to return the completed questionnaire.

parent questionnaires and centre/school records, Statistics Netherlands provided information on the country of origin of families in the home-based cohort of Pre-COOL. To analyse the migrant-native gap in achievement I distinguish between children with no migration background (both parents were born in the Netherlands) and children with a migration background (at least one parent was born outside the Netherlands). Moreover, in additional analyses I also distinguish between children from families with a Turkish and a Moroccan background.

Table 3 Classification of SES groups by parental education.

SES Category	Answer category survey	School data	Required years of education
Low (≤10)	No lbo or similar mavo	Max. lo Max. lbo	Max (mavo): 6+4=10
Medium (11-14)	havo/vwo or similar mbo or similar	Max. mbo	6+5==11 (havo); 6+6==12 (vwo) 6+4+2=12 (low level mbo); 6+4+4=14 (high level mbo)
High (≥15)	hbo	hbo/wo	6+5+4=15 (hbo); 6+6+4=16 (wo)

3.3.4 Methods

3.3.4.1 Standardisation and age corrections of achievement scores

In line with the general analytical strategy of this report I use a relative approach to measure achievement gaps. In some of the (Pre-)COOL analysis on language gaps I use composite measures, which are calculated by averaging z-scores of the relevant tests and subsequently standardising these z-score averages (as in Chapter 2 (Germany) and 5 (UK)).

In both Pre-COOL and COOL data there is substantial within-wave variation in the age of children. The aim is to remove this variation through residualisation (as in Bradbury et al. (2015); see also Chapter 2 and 6).[23] However, a problem arises when applying this approach to stages beyond kindergarten as grade retention is rather common in the Netherlands. This implies that in the primary and secondary school sample age is not only associated with test scores due to a pure age effect but also through a grade retention. In fact, whereas age is consistently positively associated with achievement until the end of kindergarten, test scores in Grade 3 and 6 generally appear to be negatively related to the child's age. These results suggest that age-variations in primary and secondary school grades capture mostly grade retention effects. Hence, residualising on age would in practice imply the exact opposite as controlling for a pure age effect.

3.3.4.2 The evolution of achievement gaps (RQ1)

To examine the extent and evolution of achievement gaps (RQ1) I estimate the relative position

[23] The main results are based on residuals from a regression of test scores predicted as a linear function of child age. Including higher order polinomials does not substantially change the results.

of the different groups using OLS, relying on the unbalanced sample and using in total 25 different dependent variables (15 tests for language and 10 for math, see Table 3). All models estimating SES effects control for migration background. The migrant-native gaps are estimated as total gaps as well as gaps net of SES. Because the trajectories are modelled for an extended observation window – from age 2 to 14 – important qualitative changes occur and (relative) gaps in achievement may not always be directly comparable over time. For example, in the primary school years reading comprehension is a relevant component of language skills, but this skill cannot be measured in the early years. The discussion of the results will therefore focus on different segments of the trajectories, mainly the early years (age 2–5) and the years from kindergarten to the end of primary school (age 5–11).

The analysis uses both cross-section weights and longitudinal (attrition) weights. Given that the unbalanced samples in COOL are representative, weights play a minor role in the models using COOL. No longitudinal weights are required for the unbalanced samples of COOL: there is panel attrition and refreshment from wave to wave and the unbalanced samples are representative. However, cross-section weights are calculated but these do not play a significant role for the results. The cross-section weights for COOL correct for the minor changes in the marginal distribution of the central variables (SES, migrant background) over time and between cohorts. For the main results, I consider the COOL3 kindergarten year 2 (age 5) – this is essentially the Pre-COOL cohort – as the base cohort and calculate weights for the older cohorts (age 8, 11, 14). More importantly, I use COOL3 distributions to calculate cross-section weights for the Pre-COOL sample. In this way, I correct for the overrepresentation of high SES in Pre-COOL. Next, to correct for panel attrition in the Pre-COOL sample, I apply inverse probability weighting. Wave 2 (age 3) of Pre-COOL is considered as the base wave: this wave has the largest number of observations as a significant number of additional children enrolled in the sampled ECEC centres after age 2. Interestingly, since age 5 data from COOL and Pre-COOL overlap, I am able to evaluate the extent of the potential remaining bias due to non-representativeness of the Pre-COOL sample. As discussed below, the Pre-COOL results are by and large consistent with the results from the representative sample, suggesting that the bias is negligible after applying the weighting strategy.

In addition to using cross-section and longitudinal weights, it is necessary to adjust standard errors for clustering given the sampling design of Pre-COOL and COOL (Abadie et al. 2017). Standard errors in the COOL analyses are clustered at school level. The standard errors in the models based on Pre-COOL are clustered at the target (Pre-)COOL school.

Finally, it should be noted that an important limitation of the approach is that I have to rely on data from older cohorts for the analysis beyond age 6: the scores for age 8 (Grade 3), age 11 (Grade 6) and Grade 14 (3rd year of secondary school) are for older birth cohorts.[24] Theoretically, age effects may be confounded with cohort effects. The results should thus be interpreted as predictions of the evolution of achievement gaps of the youngest cohort under the assumption that cohort effects do not play a major role. In Section 3.4.1 I discuss results indicating that estimates are comparable when the analysis relies on genuine cohort data (i.e. the K07/08 cohort, see Table 1).

[24] There is a cohort gap of around 9 years between the youngest cohort (Pre-COOL / COOL3-K13/14) and the oldest cohort (COOL3-K4/5), i.e. the youngest cohort was enrolled in kindergarten in 2013–2014 while the oldest cohort was enrolled in kindergarten in 2004–2005.

3.3.4.3 Explaining school gaps by preschool differences (RQ2)

To estimate the proportion of school gaps that can be attributed to achievement differences before enrolling in 1st grade, longitudinal data is required that follows children when they from preschool years through primary school. For the analysis of RQ1 I therefore use the COOL K07/08 cohort: these children were age 5 in the second year of kindergarten in the school year 2007–2008 and reached the end of primary school (6th grade) by the age of 11. The focus is on language skills (see Table 4) as these are more consistently assessed than math skills. The scores on the Language for Toddlers test, made in kindergarten, play a central role in the analysis as this test measures the initial achievement before moving to Grade 1 of primary school. Unfortunately, for the COOL K07/08 cohort no data on the Math for Toddlers is available. However, the children made an alternative test (an ordering/sorting), which I use in the empirical model (as discussed below). As I rely on the balanced sample, the number of observations is low (over 900 children) compared to the number of observations in the unbalanced. The balanced sample may not be representative if panel attrition is non-random. I correct for panel attrition using inverse probability weighting strategy. As in the models describing the evolution of gaps (RQ1), we cluster standard errors at the school level.

While the models relevant for RQ1 are estimated by OLS, an alternative estimation model is required for the analysis of RQ2. Since the analysis conditions on preschool achievement, OLS will produce biased estimates if the preschool achievement measure contains measurement error. There are several reasons why the test scores do not perfectly reflect the child's true language skills. First, the item composition of the test matters: children with the same skill level may obtain different test results as they are by chance more or less familiar with specific test items. Second, children often make guesses in tests and therefore have a non-zero probability of answering an item correctly. Third, the test day may affect scores: some students may for some random reason experience a 'bad day' and score below their potential on the test, i.e. they would have obtained a higher score if the test would have been administered on a different day. Finally, some students may exert more effort in the test than others and as a result attain a relatively high test result.

Measurement error in preschool achievement will lead to spurious regression to the mean. The intuition is that, on the one hand, children scoring in the top of the distribution are on average more likely to obtain a lower score in a subsequent measurement (i.e. some high achievers were lucky at the first test). On the other hand, children scoring in the bottom of the distribution are more likely to obtain a higher score in a subsequent measurement (i.e. some low achievers were unlucky at the first test). Moreover, it is likely that measurement error is systematically related to SES and migration background. This implies that measurement error in preschool achievement leads to an underestimation of the persistence of achievement and consequently to an underestimation of the share of school gaps that is attributable to preschool differences (and an overestimation of the additional SES effect in the school years). In line with other chapters in this report, we address this issue by applying an instrumental variable approach (see also Bradbury et al. (2015) and Jerrim & Vignoles (2013) for applications of this approach). The central idea is that in the first stage language achievement is predicted by an alternative achievement test made in the preschool years. The alternative tests used in the first stage is a kindergarten ordering/sorting test. The scores from this test are highly correlated with the kindergarten language tests. The main assumption is that the measurement error is uncorrelated between the achievement measures. This seems plausible as there is no overlap in the items and the tests are made on different days. Addressing measurement error appears to have important implications for the empirical results: in some specifications, the share of school gaps explained by preschool differences more than doubles after correcting for measurement error.

Table 4 Test data used for RQ2.

	Kindergarten (year 2) (age 5)	Grade 3 (age 8)	Grade 6 (age 11)
Test	Language for Toddlers	Language (Vocabulary; Reading comprehension)	Language (Vocabulary; Reading comprehension)

Notes: All tests used are developed by Cito.

3.4 Empirical results

3.4.1 The evolution of achievement gaps (RQ1)

3.4.1.1 Inequalities by socio-economic status

Figure 2 presents the evolution of language trajectories from age 2 to 14 by family SES, visualizing the results from 15 regression models, based on 23 tests and more than 40,000 observations. Overall, the pattern does indicate persistence or even diverging paths of different SES groups. Here I focus on three segments of the figure: the early years (age 2–5, based on the Pre-COOL test battery); kindergarten and Grade 1 (based on Pre-COOL Cito results); and kindergarten to secondary school (based on COOL3).

In the early years SES gaps in language achievement are already substantial. The high-low SES gap is around .65 SD at age 2 and increases in the following years: between age 2 and 3 the gap increases significantly to around .74. Between age 3 and 5 the high-low SES gap continues to widen at a pace of around .1 SD per year to .94. However, these increases are not statistically significant, implying that we cannot rule out stability of the SES gap in language achievement in the early years.[25] The finding that SES gaps are persistent or increase between age 2 and 5 appears not to depend on the specific language test: although gaps in vocabulary achievement are generally somewhat larger than gaps according to other language tests, the pattern over time is similar for all four Pre-COOL language tests (see Appendix 3.3).

Interestingly, the pattern of increasing gaps in vocabulary is not found when using raw (non-standardized) test scores or IRT-based test scores: in fact, these suggest convergence rather than divergence scores (see Leseman et al. 2017; Verhagen et al. 2016; see Appendix 3.1 for more details). However, as the aforementioned studies report no statistically significant change over time when using a similar indicator of SES (mother's level of education), one can argue that these results are overall consistent with the results presented in this report. Moreover, the difference can be explained by the fact that the variance of the relevant scores drops substantially from the first (age 2) until the fourth (age 5) Pre-COOL wave. As the focus of this

[25] This could be due to panel attrition of more than two-thirds of the sample from age 3 to 4 when children enrol in kindergarten. Given the relatively small sample size, the power of test is rather low.

report is on the development of relative achievement gaps, I adjust for changes in the variance of the test score distribution. These results highlight that absolute and relative gaps should not be confused.

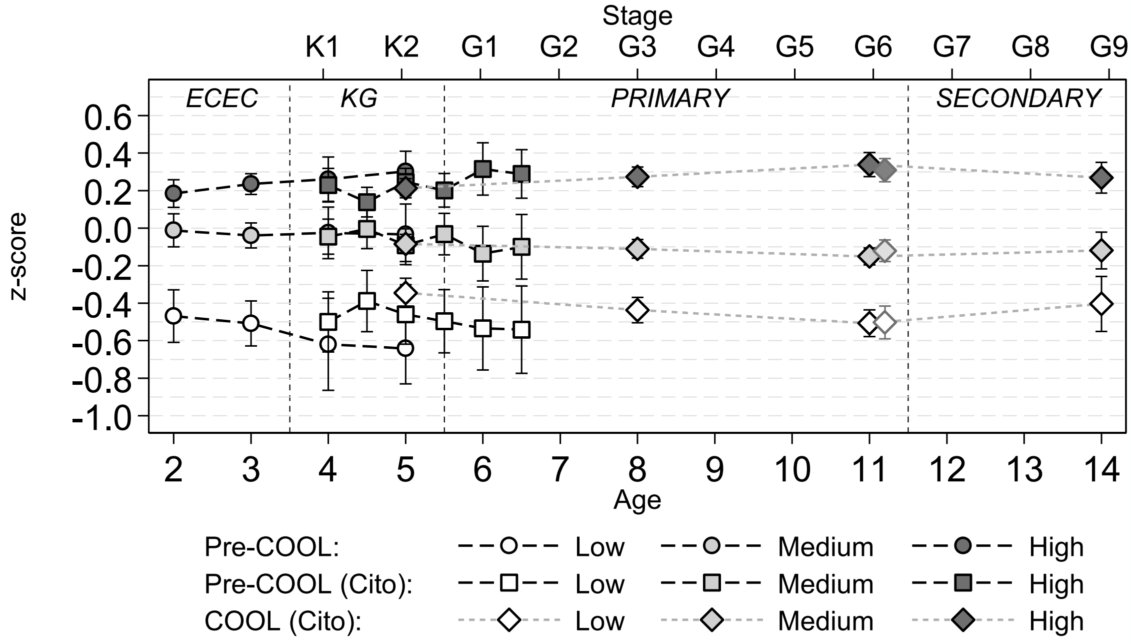


Figure 2 Language z-scores of children by SES (parental education).

Notes: Predictions are from stage-specific regression models. Grey-bordered diamonds represent estimates based on Cito End test (language) scores (age 11). Long-dashed black lines connect data within the same cohort: results before age 7 are based on the same cohort; results beyond age 7 are based on older cohorts. Capped spikes indicate 95% confidence intervals. Stage: K = Kindergarten, G = Grade level in school.

The findings concerning the Pre-COOL Cito language test indicate persistence of the gap in the first years after kindergarten and school enrolment. The analysis based on the largest sample (age 5; mid-year 2 of kindergarten) indicates a high-low SES gap of around .7 and a medium-low SES gap of about half that size. Compared to the Pre-COOL test battery results, the gaps based on this test are somewhat smaller. Nevertheless, the estimates for high and medium SES children are almost identical when considering the same time periods. For the low SES group, there is a visible but small difference and the confidence intervals overlap. Interestingly, the Pre-COOL Cito and and COOL Cito results, both based on the Language for Toddlers test, are almost identical. However, the Pre-COOL low SES point estimate is slightly lower than the COOL counterpart. This suggests that the Pre-COOL results are highly comparable to those based on the representative COOL sample. However, low SES may be more negatively selected into the Pre-COOL sample, implying that models based on Pre-COOL probably somewhat overestimate SES language gaps in the population.

The third relevant segment of Figure 2 describes the evolution of SES gaps in language achievement when children move from kindergarten through primary school. Again, the estimates point out divergence of trajectories, although high-low SES gaps grow at a slower pace in this phase than in early childhood: from .56 SD in kindergarten to over .85 by the end of primary school (i.e. around half a SD per year). Yet, low and medium SES children follow parallel pathways, indicating that the main development over time can be characterized as high SES children moving away from the rest of the population. Interestingly, the results based on the language component of the Cito End test are highly consistent with our main end of primary school estimates (the former indicates a .81 low-high SES gap).

Next, Figure 2 does indicate a pattern of convergence after leaving primary school. However, one may argue that this result can be explained by the use of a different composite language measure (vocabulary and reading comprehension versus grammar/orthography and reading comprehension). Nevertheless, even when I use tests measuring the same domain (i.e. test data on reading comprehension is available for both stages), I find a reduction of the SES gap after enrolling in secondary school (from around .8 to .7 SD). As mentioned earlier, it should be stressed that the secondary school sample is not fully representative, and this result should therefore be interpreted cautiously.

Finally, it should be noted that the results beyond age 7 are all based on COOL3. This dataset concerns older cohorts – e.g. children in the Grade 6 sample were born around 6 years earlier than the children in the sample used to describe the development of gaps in the early years. In additional analyses I have used data representing a single cohort (using all three waves of COOL) to estimate the age 5-11 trajectories. The results can be found in Appendix 3.2. The estimates clearly show that there are hardly any cohort differences in stage-specific positions and trajectories. Although the future development of the trajectories of the Pre-COOL cohort remains of course uncertain to some extent, these results indicate that cohort effects are probably negligible when the birthyears from the different cohorts are not too far apart.

While Figure 2 provides evidence on the evolution of language trajectories, Figure 3 represents the results of an equivalent analysis for math achievement scores. The figure represents the results from 10 regressions using in total over 37.000 observations on the development of math trajectories by SES from age 3 to 14. As in language domain, math gaps are already substantial in the early years before entry into kindergarten. The patterns for math indicate a larger degree of stability: between age 3 and 5, the high-low gap hovers around .7; and the medium-low gap stays close to half of that gap. Although the age 3 gap in math is similar to the age 3 gap in language achievement, the math gap is more persistent in the early years. Moreover, between kindergarten and the end of primary school SES trajectories diverge to some extent: the high-low SES gap increases from .56 to .73 SD. Compared to the results for language, this change is rather small (i.e. less than .03 SD per year).[26] The gap remains very stable as children leave primary school and move through the first years of secondary school. Again, there appears to be hardly any difference between the age 5 Pre-COOL and COOL results, suggesting that the Pre-COOL results are representative.[27] Finally, additional results also indicate negligible cohort differences in math trajectories, which once more indicates that the presented results beyond age 6 are plausible predictions of the development of the Pre-COOL cohort.

---

[26] Results from pooled regressions show that this increase in the SES math gap is not statistically significant.

[27] As mentioned above, the age 5 test data are not perfectly comparable between COOL1 and COOL3 as a different test was used.

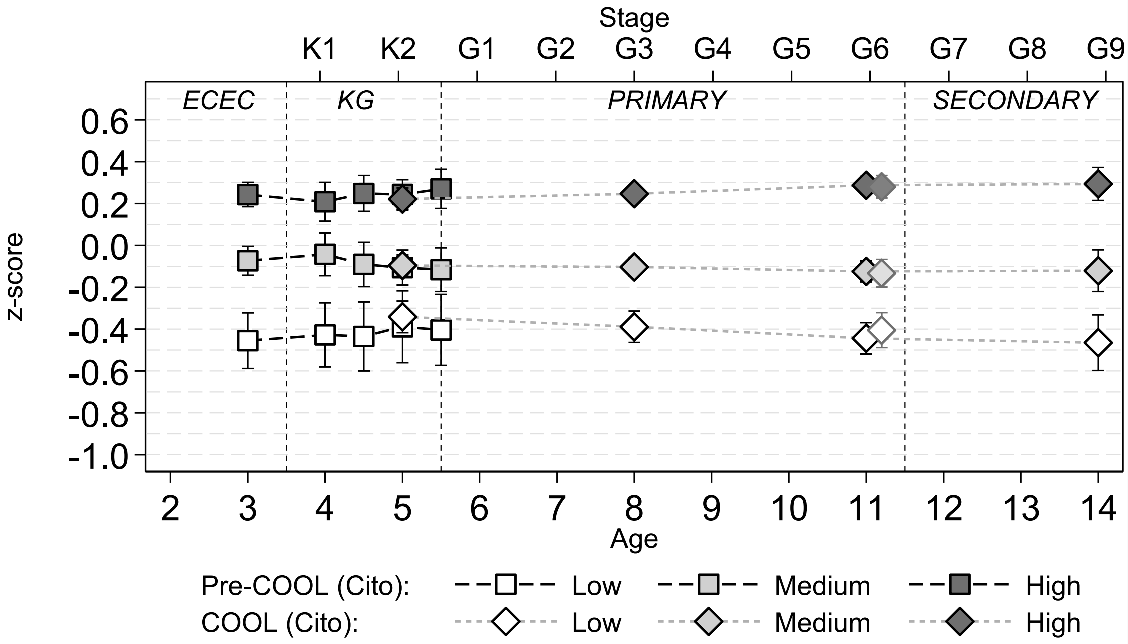


Figure 3 Math z-scores of children by SES (parental education).

Notes: Predictions are from stage-specific regression models. Grey-bordered diamonds represent estimates based on Cito End test (math) scores (age 11). Long-dashed black lines connect data within the same cohort: results before age 6 are based on the same cohort; results beyond age 7 are based on older cohorts. Capped spikes indicate 95% confidence intervals. Stage: K = Kindergarten, G = Grade level in school.

3.4.1.2 Inequalities by migration background

Figure 4 plots the results from models estimating migrant-native gaps in language achievement from age 2 to 14. Overall, the figure indicates that gaps close, especially in early childhood. As in the discussion on SES trajectories, I discuss three separate segments of the figure. First of all, the results from Pre-COOL show that language achievement differences between children with and children without a migration background are sizeable before children enrol in kindergarten (.85–.9 SD). SES differences can only account for a small share of this gap: after controlling for SES, the migrant-native gap at age 2–3 remains substantial (around .77 SD). However, more than half of this gap disappears after children move into kindergarten. Next, estimates based on the Pre-COOL Cito language test show that migrant-native gaps in kindergarten are substantial in kindergarten (.48–.59 SD for mid-year 2 kindergarten, depending on whether SES differences are accounted for).[28] However, the figure also shows that gaps continue to narrow during the period in kindergarten.

Furthermore, the total (net of SES) gaps decline during the primary school phase by 18% (38). Estimates based on the Cito End test data point out an even more dramatic narrowing of the gap: the total (net of SES) language gap reduces by about 50% (75). No significant migrant-native gaps are found for children in year 3 of secondary school, suggesting a further convergence of trajectories after the transition to secondary school. Hence, the estimation results from the three separate segments of Figure 4 consistently point out that the migrant-native gap declines

[28] The kindergarten results from Pre-COOL Cito are again highly consistent with the COOL results.

between age 2 and 14. While the migrant-native gap is large before school entry, it appears that migrants catch up in primary and secondary school.

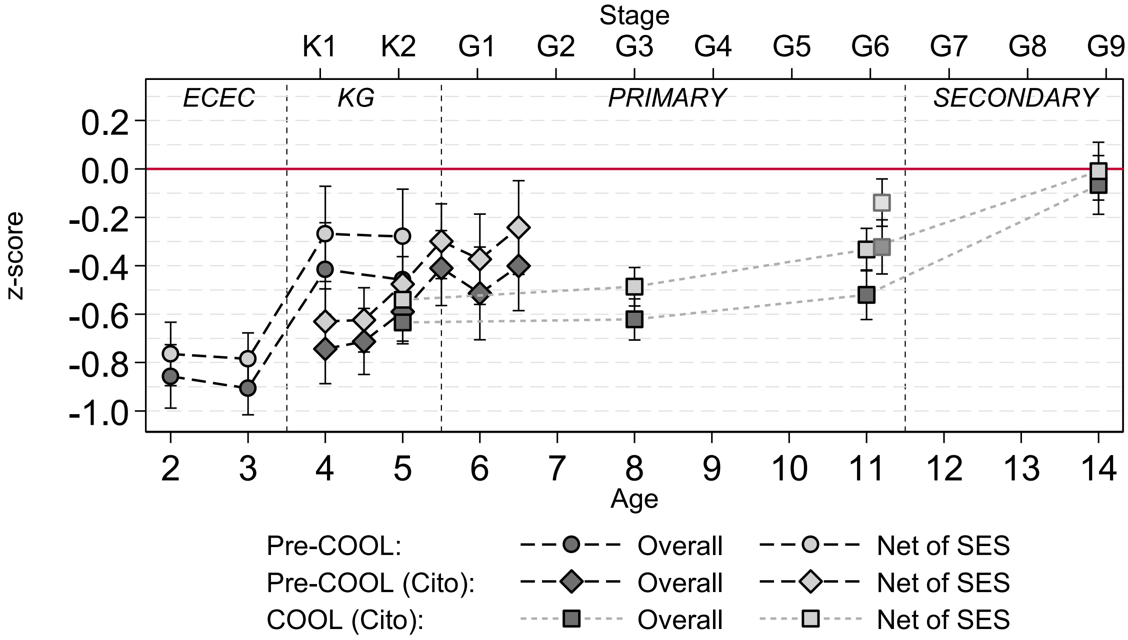


Figure 4 Total and net (SES) migrant-native gaps in language achievement.

Notes: Predictions are from stage-specific regression models. Grey-bordered diamonds represent estimates based on Cito End test (math) scores (age 11). Long-dashed black lines connect data within the same cohort: results before age 7 are based on the same cohort; results beyond age 7 are based on older cohorts. Capped spikes indicate 95% confidence intervals. Stage: K = Kindergarten, G = Grade level in school.

Although migrants appear to catch up in general, Figure 5 (total gap) and 6 (gap net of SES) indicate that the extent and evolution of language achievement gaps depend on the country of origin of the child's parents. The results clearly show that children with a Turkish background consistently and considerably lag behind.[29] Children with a Turkish background consistently score lower than children with a Moroccan background and other migrant groups. The total Turkish-native gap in language is especially large in the early years (1.5–1.9 SD at age 2–3), but the gap appears to narrow in the early years. However, the results based on COOL indicate that the Turkish-native gap in language achievement is also large in kindergarten: almost 1 SD (.8 SD net of SES). Moreover, children with a Turkish background show no signs of catching-up in language as the total gap hovers around 1 SD in Grade 3 and Grade 6.[30] Although less pronounced, language achievement differences between children from Moroccan families and children with native Dutch parents are also large in the early years (over 1 SD at age 2–3). However, for this group the migrant-native gap narrows somewhat during the school years; this catching-up pattern is stronger when conditioning on SES.

---

[29] Note that the Pre-COOL results are based on a relatively small number of observations; migrant group-specific results should therefore be interpreted cautiously. Results from COOL (5-11) are based on much larger samples and are therefore more reliable.

[30] Given that the COOL secondary school sample is not fully representative and children with a Moroccan and Turkish background are underrepresented, no secondary school gaps are presented.

Overall, from age 2 until the end of primary school, children with a Moroccan and Turkish background are in a disadvantaged position. Although these groups catch up to some extent in the preschool and kindergarten stage, the language achievement penalties appear to be rather persistent once children enter primary school. The results concerning the Turkish-native gap are striking as they point out large gaps in the preschool period and almost no closing of the gap in the primary school phase.

Do migrant-native gaps in math achievement evolve according to similar patterns? Figure 7 shows the evolution of migrant-native gaps in math achievement over the educational career. While the migrant-native gap in math is sizeable and significant in the early years (the age 3 total gap is .56 SD; the net of SES gap .43 SD), the gap in math is consistently smaller than the migrant-native gap in language (Figure 4). Moreover, the evidence points out that migrants catch up in math during primary school: by the end of primary school the migrant-native gap net of SES is small (less than .1 SD) and only marginally significant ($p < .05$). Results from the math component of the Grade 6 Cito End test in fact indicate no significant difference in math achievement scores. Interestingly, the results show an (insignificant) migrant premium for this test when conditioning on SES. This evidence is important from a life course perspective, as Cito End test scores are relevant for the track placement in secondary school.

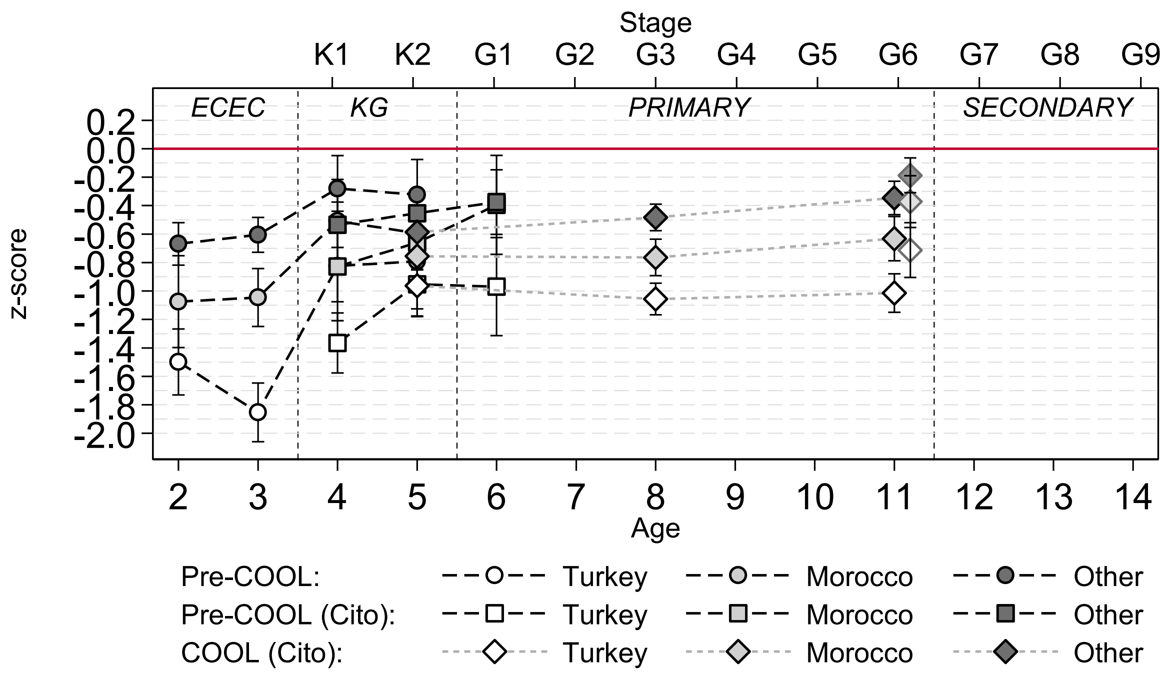


Figure 5 Total migrant-native gaps in language achievement by country of origin.

Notes: Predictions are from stage-specific regression models. Grey-bordered diamonds represent estimates based on Cito End test (math) scores (age 11). Long-dashed black lines connect data within the same cohort: results before age 7 are based on the same cohort; results beyond age 7 are based on older cohorts. Capped spikes indicate 95% confidence intervals. Stage: K = Kindergarten, G = Grade level in school.

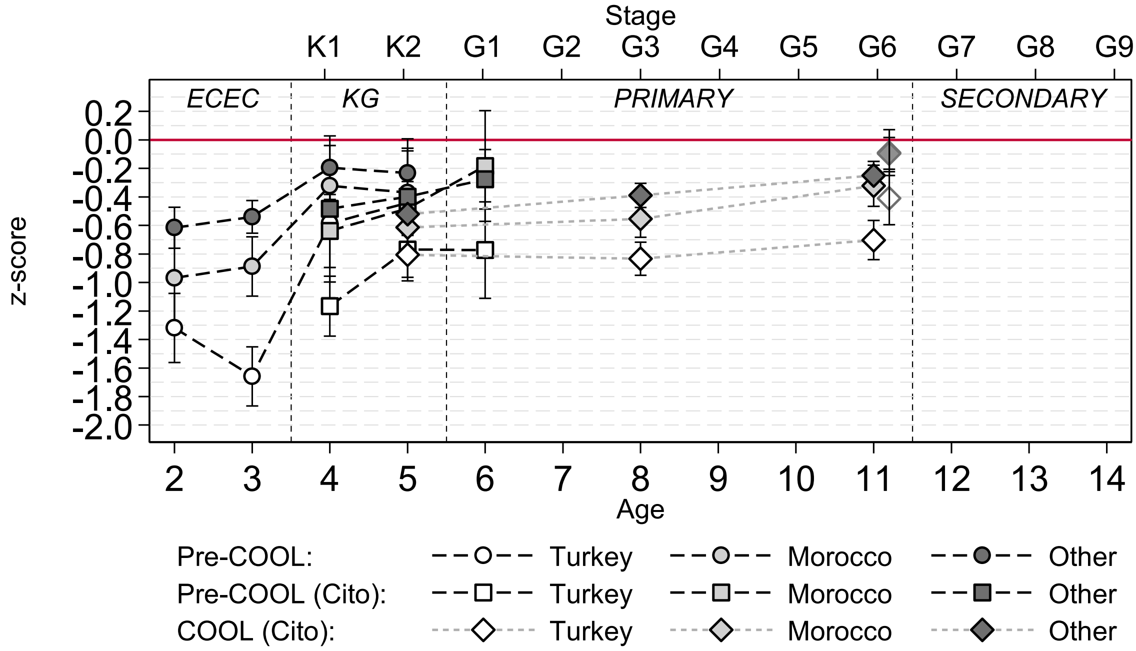


Figure 6 Net (of SES) migrant-native gaps in language achievement by country of origin.

Notes: Predictions are from stage-specific regression models. Grey-bordered diamonds represent estimates based on Cito End test (math) scores (age 11). Long-dashed black lines connect data within the same cohort: results before age 7 are based on the same cohort; results beyond age 7 are based on older cohorts. Capped spikes indicate 95% confidence intervals. Stage: K = Kindergarten, G = Grade level in school.

Figure 8 and 9 present the estimation results of math gaps by the country of origin of the child's parents. As in the language domain, children with a Turkish background perform worse in math than natives and other migrant groups in the early years: at age 3, the total Turkish-native gap in math achievement is around 1 SD. Nevertheless, the evolution of math gaps differs from the evolution of language gaps. In general, Turkish and Moroccan children catch up significantly during the primary school period (e.g. the Turkish-native gap decreases by around two thirds). It is striking that these two groups catch up more rapidly than other migrant groups. While in the preschool period other migrant groups on average outperform Turkish and Moroccan children, the results from models conditioning on SES point out that by the end of primary school the gaps are actually smaller for Turkish and Moroccan children. Consequently, no significant Turkish-native and Moroccan-native achievement gaps in math are found in Grade 6 when controlling for SES differences. The results based on the math component of the End Cito test show that children with a Turkish background score significantly lower than natives, although this gap is relatively small (.18 SD). However, for this test the Moroccan-native gap is not significant. Moreover, when conditioning on SES, the sign of the gap reverses, indicating an insignificant Turkish and a significant (.2 SD) Moroccan math premium.

All in all, the migrant-native gap in math scores are substantial in the early years but narrow considerably during primary school. The evolution of math trajectories of Turkish and Moroccan children is striking as they catch-up more than other migrants and the gap observed by the end of primary school is rather small and can be fully explained by SES.

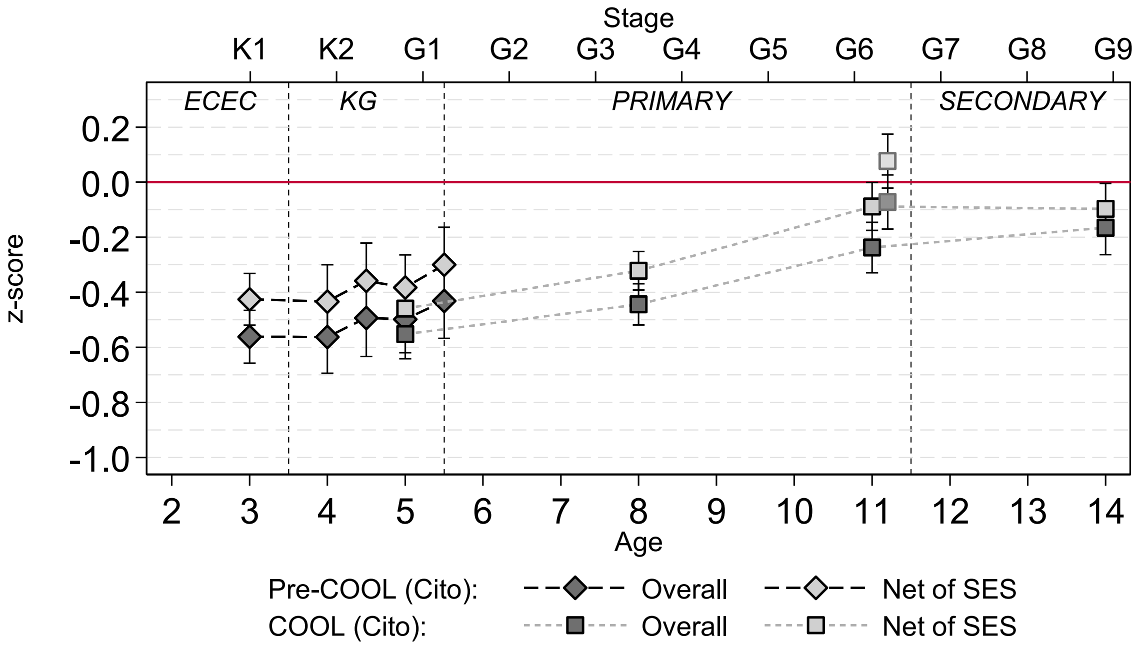


Figure 7 Total and net (SES) migrant-native gaps in math achievement.

Notes: Predictions are from stage-specific regression models. Grey-bordered diamonds represent estimates based on Cito End test (math) scores (age 11). Long-dashed black lines connect data within the same cohort: results before age 6 are based on the same cohort; results beyond age 7 are based on older cohorts. Capped spikes indicate 95% confidence intervals. Stage: K = Kindergarten, G = Grade level in school.

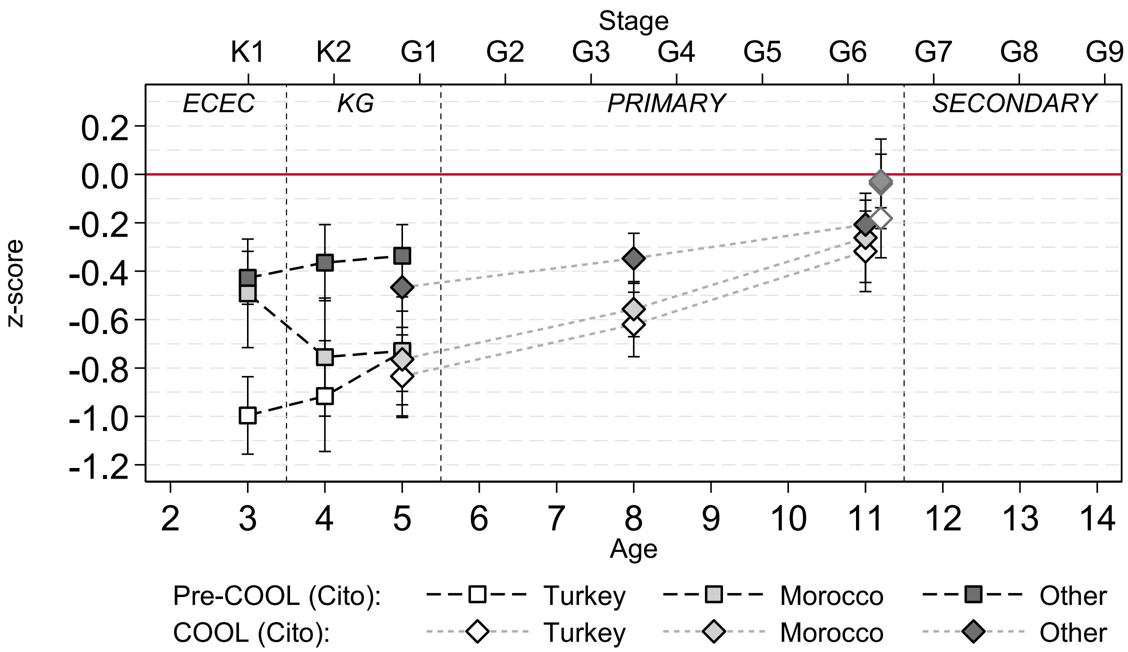


Figure 8 Total migrant-native gaps in math achievement by country of origin.

Notes: Predictions are from stage-specific regression models. Grey-bordered diamonds represent estimates based on Cito End test (math) scores (age 11). Long-dashed black lines connect data within the same cohort: results before age 6 are based on the same cohort; results beyond age 7 are based on older cohorts. Capped spikes indicate 95% confidence intervals. Stage: K = Kindergarten, G = Grade level in school.

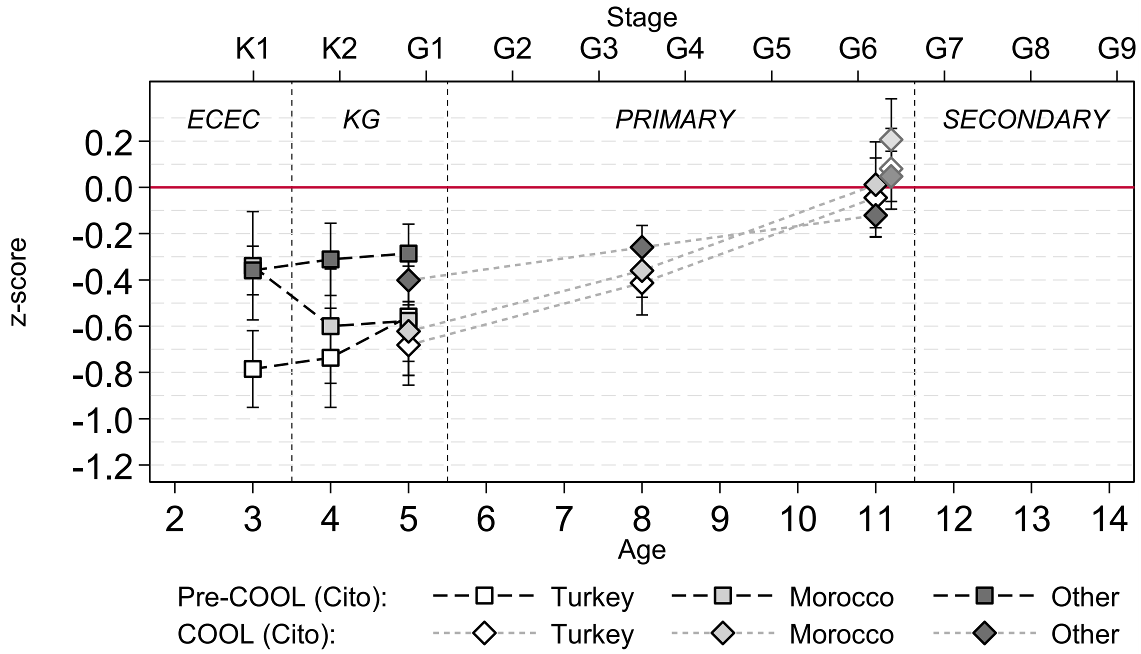


Figure 9 Net (of SES) migrant-native gaps in math achievement by country of origin.

Notes: Predictions are from stage-specific regression models. Grey-bordered diamonds represent estimates based on Cito End test (math) scores (age 11). Long-dashed black lines connect data within the same cohort: results before age 6 are based on the same cohort; results beyond age 7 are based on older cohorts. Capped spikes indicate 95% confidence intervals. Stage: K = Kindergarten, G = Grade level in school.

3.4.2 Explaining school gaps by preschool differences (RQ2)

3.4.2.1 Inequalities by socio-economic status

The proportion of school gaps attributable to preschool differences

The results presented in Section 3.4.1.1 provide evidence that SES gaps widen as children move through school. To examine the extent to which social inequalities in educational achievement in school can be explained by preschool inequalities (RQ2) I compare aggregate SES gaps with gaps according to common trajectories. Figure 10 presents average (observed) and (predicted) common trajectories (CT) of children from low, medium and high SES families. The average trajectories are based on a balanced sample, using COOL1–3. In general, these results are similar to the results presented in Figure 2 (unbalanced sample, using COOL3) and in Appendix 3.2 (unbalanced sample, using COOL1–3). As described in Section 3.4.1.1 for the unbalanced sample, in the balanced sample SES gaps also increase during the primary school years. This appears to be mainly the result of high SES children moving away from the rest of the population: the high-low gap increases from around .5 in kindergarten to around .8 in the final grade of primary school.

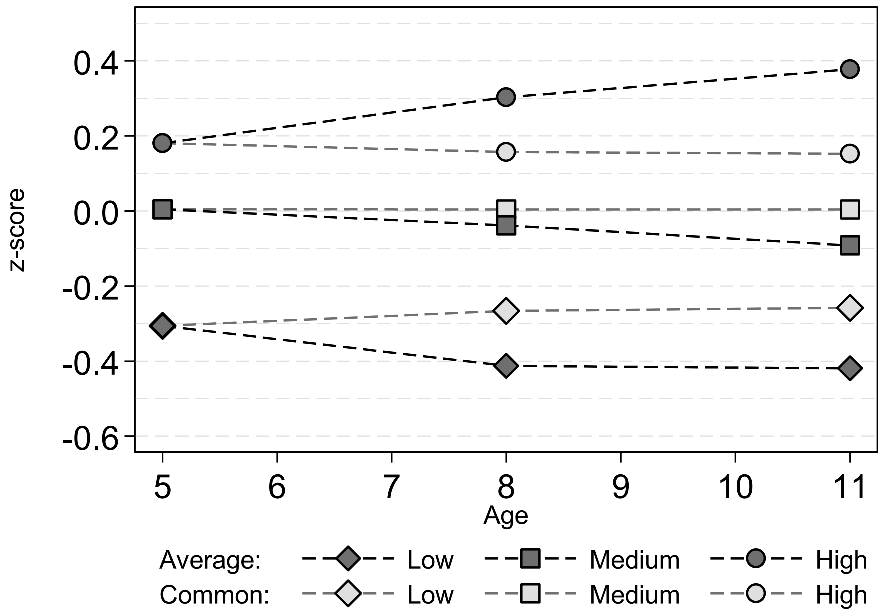


Figure 10 Aggregate and common trajectories by SES.

The common trajectories shows the predicted path of different SES groups under the assumption that all three SES groups would have developed in the same way after kindergarten. As discussed above, these IV estimates are correcting for measurement error. The common trajectories represent a natural counterfactual condition of no SES effects in primary school. Given regression to the mean (i.e. due to transitory shocks), one can expect that the end of school gaps are smaller according to the common trajectory model than the actual preschool gaps. This implies that if SES had no additional effect in primary school, the aggregate gaps in school can be expected to be smaller than the preschool gaps.

Whereas the aggregate SES gaps in the final grade of primary school are .33 SD (medium-low) and almost .8 SD (high-low), these gaps are predicted to be .26 and .41 respectively in the absence of additional SES effects. This suggests that 80% of the medium-low gap and 52% of the high-low gap can be attributed to differences in preschool achievement. Hence, comparing the aggregate and common trajectories, one can conclude that the majority of language achievement gaps observed in the final grade of primary school can be explained by preschool inequalities.

Instead of comparing aggregate with common trajectories, I can estimate SES gaps in language achievement in primary school while conditioning on preschool achievement levels (again using an IV approach): see Table 5 for the estimation results. The findings show that achievement scores are highly persistent: a 1 SD increase in preschool scores is associated with a .8 SD increase in scores obtained in Grade 6. When comparing aggregate gaps with gaps conditioning on preschool achievement level, it appears that 75% of the medium-low and 48% of the high-low gap can be explained by differences in preschool achievement. Overall, the estimated percentage explained by preschool achievement is comparable (though slightly lower) with the estimates based on the aggregate-common trajectory comparisons discussed above.

Table 5 SES gaps in language achievement and the role of preschool differences (RQ2).

		Stage (approx. age)			
		Grade 3 (age 8)		Grade 6 (age 11)	
		OLS	IV	OLS	IV
SES (ref. low)					
	Medium	0.374**	0.116	0.327**	0.081
		(0.146)	(0.139)	(0.141)	(0.164)
	High	0.716***	0.311**	0.797***	0.411**
		(0.193)	(0.147)	(0.191)	(0.178)
Language score KG			0.832***		0.792***
			(0.097)		(0.107)
N		870	870	870	870
Low-Medium gap					
% preschool			*69*		*75*
% additional			*31*		*25*
Low-High gap					
% preschool			*57*		*48*
% additional			*43*		*52*

Notes: Clustered standard errors in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

Diverging trajectories

While the evidence indicates that preschool achievement gaps are rather persistent, there appear to be significant additional SES effects beyond the preschool years. I examine whether these additional SES effects depend on the child's position in the kindergarten achievement distribution: are high SES children more likely to recover from a bad start in kindergarten than low SES children? Does the position of low SES children who are low achievers at age 5 worsen or improve during the primary school years? The predictions based on IV models allowing for the interaction between SES and age 5 achievement level are presented in Figure 11. The figure shows predicted trajectories of low, medium and high SES children who performed low (–1 SD above the average), average or high (+1 SD above the average) in a language test in kindergarten. It should be noted that these results should be interpreted as simulated rather than actual trajectories and that the point estimates are surrounded by large confidence intervals.

The figure clearly shows that achievement of high SES children systematically – i.e. independent of the initial position – improves more than the position of low and medium SES children. The predictions also suggest that low–achieving high SES recover to a large degree from their initial disadvantage between kindergarten and 6th grade (from –1 SD to –.6 SD). The position of medium SES remains quite stable during the school years and this holds for low, average and high achievers.

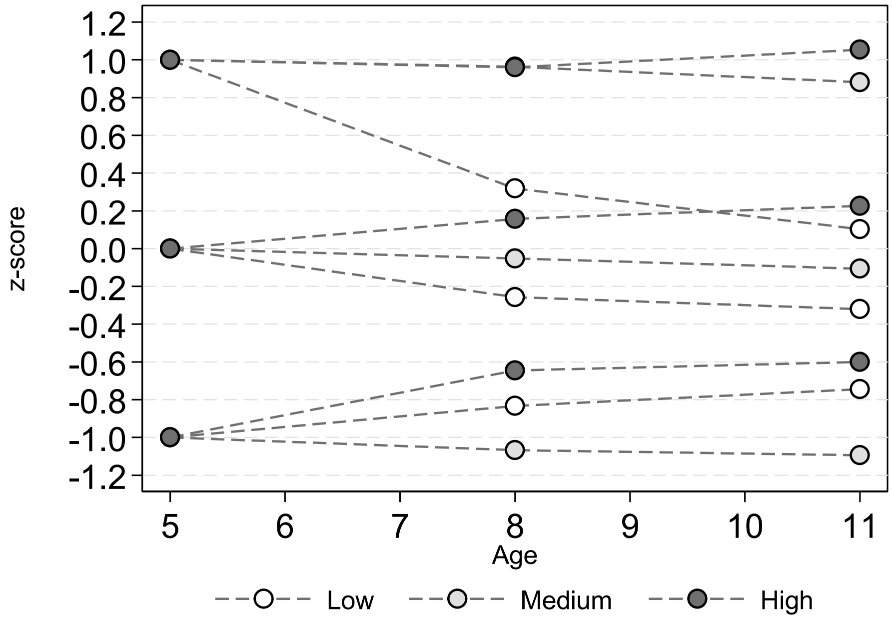


Figure 11 Diverging trajectories by SES.

The most remarkable results concern low SES children in the bottom and top of the kindergarten achievement distribution. The figure shows that low achieving, low SES children are not being left behind but catch-up somewhat (i.e. by more than .2 SD) during the primary school period. The point estimates actually indicate that this group outperforms low achieving, medium SES children. Nevertheless, regarding children in the top of the kindergarten achievement distribution, low SES children appear to lag behind their medium and high SES peers. Particularly among high achievers in kindergarten, the increase in the SES gap is substantial: by the end of primary school, the position of low SES children starting in the top of the achievement distribution is overtaken by high SES children with an average kindergarten achievement level. In general, these estimates suggest that achievement is less persistent among low SES children – i.e. they have a stronger tendency to move towards the mean of the achievement distribution. Although we cannot identify the factors that cause these patterns, two important features of the Dutch system may provide an explanation for the striking results. First, school segregation by socio-economic background is relatively high in the Netherlands, which may lead to downward mobility of high-achieving, low SES children. Second, schools with high concentrations of disadvantaged children receive additional funding to improve the overall quality, which may boost upward mobility of low-achieving, low SES children.

Implications for secondary school track placement

The maximum longitudinal component of COOL is 6 years, implying the data can be used to follow children from preschool to Grade 6 (the end of primary school), but not into secondary school. The role of preschool achievement in secondary school track enrolment can therefore not directly be tested. However, the data does contain information on variables that determine track placement in secondary school. First, track placement is based on the results of a standardized test. In the relevant time period, around 85% children participated in the Cito End test. The standardized test scores correspond to a specific track recommendation (the “test track recommendation”). Second, the school recommendation plays an important role. School recommendations take into account achievement scores but also other characteristics of the child such as motivation and attitudes towards school.

Appendix 3.4 replicates the main findings presented in Table 5 using Grade 6 Cito End test results. Preschool language scores (net of SES) appear to be strong predictors of these test results: a 1 SD increase in preschool achievement is associated with a .7 SD increase in the Cito End language score and a .9 SD increase in the Cito total score. As a result, the test track recommendation is also significantly predicted by preschool achievement: a 1 SD increase in preschool achievement increases the probability to receive a high test track recommendation by over 30% (see Appendix 3.4).

Consistent with many other studies (e.g. Inspectorate of Education, 2018), significant SES effects on Cito End test scores and track recommendations are found. The SES effect on the test track recommendation is slightly smaller than the SES effect on the school track recommendation. The results point out that a sizeable share (30–50%) of the SES effect on Cito End test scores is explained by differences in preschool scores. Consequently, preschool language achievement explains a substantial proportion of the SES effect on the probability to receive a high track recommendation (defined as a havo or vwo track recommendation): 29 (high-low gap) to 67% (medium-low gap). Similarly, the SES effect on the school track recommendation can, to a significant degree, be attributed to preschool differences in language achievement: 27 (high-low gap) to 51% (medium-low gap). Hence, the results imply that SES inequalities in the two central dimensions determining track placement can be partially explained by achievement differences in kindergarten.

3.4.2.2 Inequalities by migration background

The proportion of school gaps attributable to preschool differences

As discussed in Section 3.4.1.2, the evolution of aggregate migrant-native trajectories in language achievement shows some signs of convergence: gaps appear to narrow as children move through primary school. Despite the process of convergence, migrant-native gaps remain sizeable and significant even in the last grade of primary school. To estimate the share of migrant-native school gaps that can be attributed to preschool differences, we follow an equivalent approach as in 3.4.2.1. First, I compare aggregate and common trajectories (see Figure 12). The common trajectories represent the counterfactual trajectories of no additional migrant effects in the primary school period. According to this approach, the migrant-native gaps by the end of primary school can be entirely explained by preschool differences if actual gaps by the end of primary school would be identical to those predicted by the common trajectory models. The results show that the total migrant-native gap according to the common trajectories would decline from .85 in kindergarten to .68 in Grade 6 (Figure 12, left panel). In the absence of additional migrant penalties during school the achievement gap would narrow as a result of regression to the mean (note that the IV approach controls for “spurious” regression to the mean due to measurement error). Nevertheless, the reduction of the aggregate total migrant-native gap appears to be even larger than predicted by the common trajectories model. This pattern is also found when conditioning on SES differences (Figure 12, right panel). Hence, the aggregate-common trajectory comparison implies that differences in preschool achievement fully account for the migrant-native gap in school.

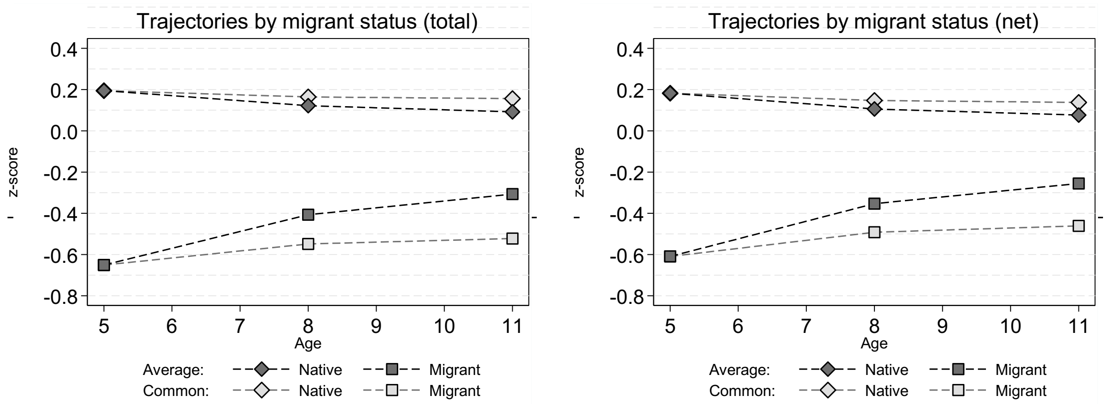


Figure 12 Aggregate and common trajectories by migration background.

Table 6 presents the results of IV models estimating migrant-native gaps in language achievement in primary school. Whereas aggregate gaps are significant in Grade 6 (total: .4 SD; net of SES: .33 SD), the coefficient indicating the migrant-native gap becomes positive (though generally not significant) when controlling for differences in preschool achievement. This means that migrants entering school with the same language achievement as their native counterparts would experience a migrant premium rather than a penalty. Again, this result points out that migrants catch up during the primary school phase and that migrant-native gaps in language achievement can be entirely explained by preschool differences.

Table 6 Migrant-native gaps in language achievement and the role of preschool differences.

	Stage (approx. age)			
	Grade 3 (age 8)		Grade 6 (age 11)	
	OLS	IV	OLS	IV
Panel A: Overall gaps				
Migrant (ref. native)	-0.529***	0.207	-0.399**	0.315
	(0.156)	(0.173)	(0.151)	(0.225)
Language score KG		0.870***		0.844***
		(0.102)		(0.111)
N	870	870	870	870
Panel B: Gaps net of SES				
Migrant (ref. native)	-0.459***	0.200	-0.332**	0.296
	(0.160)	(0.156)	(0.139)	(0.198)
Language score KG		0.832***		0.792***
		(0.097)		(0.107)
N	870	870	870	870

Notes: Clustered standard errors in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

Implications for secondary school track placement

On average, migrants obtain lower primary school achievement scores. Results based on COOL also show that migrants perform worse on the Cito End test (Appendix 3.5). The migrant-native gap is rather small and not always statistically significant. Given that secondary school recommendation is to a large extent based on achievement scores, a migrant penalty on secondary school recommendation can be expected. The estimation results indeed show that this is the case (see Appendix 3.5): migrants receive a lower test track recommendation (because they perform worse on the Cito End test) as well as a lower school track recommendation. Compared to natives, migrants face a 23 (test) and 16% (school) lower probability to receive a high track recommendation. Conditioning on SES reduces this gap substantially though. Furthermore, the point estimates indicate that migrants are more likely than natives to receive a high track recommendation when controlling for preschool differences in language achievement. Interestingly, this migrant premium is stronger for the school than for the test track recommendation. A potential explanation is that school track recommendations are not only based on past and current achievement levels, but also take into account future improvements of children's achievement levels.

3.5 Conclusions

This chapter combines data from two related longitudinal studies – COOL and Pre-COOL – to examine the early roots of social and migration-related inequalities in the Netherlands. The findings reported in the chapter document how language and math gaps evolve from early childhood (age 2–3) to adolescence (age 14). During this extensive observation window, children make several important life transitions and experience important qualitative changes in development.

The evidence clearly shows that SES gaps in language and math are already large before children enrol in kindergarten. SES gaps in language appear to be larger than in math though. The analysis of the evolution of gaps over an extended observation window (age 2–14) show that SES gaps are persistent or widen somewhat during the school years. In general, divergence is more pronounced in the language domain. The evidence indicates that a substantial share of language achievement gaps in school can be attributed to differences in preschool achievement. Concerning gaps measured in the final grade of primary school (Grade 6), 75 to 80% of the medium-low SES gap can be explained by preschool differences in achievement. A smaller share of the high-low SES gap is attributable to preschool differences (around 50%), which can be explained by the result that high SES children appear to move away from the rest of the population. Overall, the findings presented in this chapter imply that at least half of the SES gaps in school can be explained by differences that were already present before school entry.

Interestingly, simulated trajectories indicate that, depending on their preschool achievement, low SES children experience either a rise or fall of their relative position as they move through primary school. On the one hand, low SES children starting in the bottom of the achievement distribution preschool appear to catch up more than their medium SES counterparts and nearly as much as their high SES counterparts. On the other hand, low SES children who are in the top of the achievement distribution before school entry almost completely lose their initial advantage by the time they reach the final grade of primary school. The high degree of both downward and upward among low SES children may be due to the Dutch 'high segregation, high compensation' context: while school segregation is relatively high in the Netherlands, schools with a higher concentration of disadvantaged children receive significant amounts of additional funding.

The chapter also provides evidence of significant migrant-native gaps, which are particularly pronounced in the early years. As for SES inequalities, migrant-native gaps are larger in language than in math achievement. The extended observation window (age 2–14) results show a striking pattern of a narrowing migrant-native gap. Migrant and native trajectories converge more in the math domain; language gaps are overall more persistent. In general, there is strong evidence pointing out that migrants catch up in the early years and during primary school. The findings indicate that the migrant-native gap observed in primary school can be fully attributed to preschool differences in achievement. Compared to natives with the same preschool language achievement, migrants obtain an (insignificant) achievement premium.

Focusing on different groups of migrants, the results show that Moroccan and Turkish migrants perform worse than natives and other migrants in both domains. During the entire observation window, Turkish migrants consistently appear to be the most disadvantaged group. Moroccan children catch up somewhat in the language domain, but Turkish children do not seem to recover from their initial disadvantaged position in language achievement during the school years. However, there is compelling evidence indicating that both groups catch-up almost completely in the math domain.

There is now mounting evidence that the early years are important for later school achievement, educational attainment, labour market success and other relevant life outcomes (e.g. Doyle et al. 2009; Heckman & Mosso 2014; Kautz et al. 2014). This chapter further quantifies the relevance of the early years by describing the evolution of gaps from early childhood to adolescence and by estimating the role of preschool differences for later school achievement levels. The empirical results presented here once more highlight that the early years matter. Targeted high-quality early childhood education and care programs have the potential to significantly reduce achievement disadvantages in the early years (Elango et al. 2015). Moreover, since universal (high-quality) early childhood education and care programs often generate significant benefits only for lower SES children (van Huizen & Plantenga 2018), universal programs have also the potential to level the playing field. Although increased investments in such programs will involve substantial public short-term costs, various studies indicate that the societal long-run benefits will probably outweigh these costs (Elango et al. 2015; van Huizen et al. 2018). The reported evidence provides a strong rationale to prioritize programs and policies that effectively reduce achievement disadvantages in the early years.

4 NORWAY

Socio-economic and Migration-related Inequality in Early Language Development in Norway[1]

Henrik Daae Zachrisson♠ & Luisa Ribeiro♠

♠University of Oslo

4.1 Background and institutional context

Norway is a wealthy social democracy with low unemployment and low economic inequality, at least relative to other wealthy nations (Eurofound 2017). Nonetheless, the distribution of household income is highly skewed in Norway: the top 10% of households own 53% of Norway's wealth and the top 1% own 21% of the wealth (Norway 2015). Child poverty rates (calculated as the proportion of children living in a family with income under the EU 60% poverty line for three consecutive years) surged from 4.1% in 1997–1999 to 10% in 2013–2015 and are projected to reach 15% in ten years (The Norwegian Directorate for Children 2017).

Norway provides free health care for all, including free access to maternal health and well-baby clinics, and regular health check-ups prior to school age, when the school health services take over. Norway also provides universal and subsidized ECEC from age 1, following one-year paid parental leave. ECEC for 1 and 2-year-olds was expanded considerably during the 2000s, increasing the national coverage from about 30% to 80% over a ten-year period. The percentage of children aged 1–2 years in ECEC is currently around 83%, whereas the percentage of 3/5-year-old children is nearly 97% (Norway 2018). Parent fees were capped to about EUR 250/month in 2005, with a sliding scale for lower income parents, and have remained at this level since. ECEC is by law an educational enterprise, and ECEC centres are required to follow national curriculum guidelines (for an elaboration of the ECEC expansion in Norway, see Dearing, Zachrisson, Mykletun, and Toppelberg, 2018). Centres may be publicly or privately owned, but all ECEC centres receive public subsidies and are required to have the same quality standard. There is a recommended adult-child ratio of 3:10 for children under 3 and 3:19 for children older than 3, while a minimum teacher-child ratio of 1:10 and 1:19 is required for children under and over 3, respectively. Teachers are required to have at least a 3-year degree in early childhood education. Although not yet met by all centres, there is a relatively homogenous quality standard across most centres (Winsvold & Guldbrandsen 2009).

Achievement gaps between the most and least affluent students in PISA (2011) are somewhat smaller in Norway (75% of a SD) compared to, for example, the US (1.25% of a SD). Importantly, gap sizes are strongly related (r=.64) to national levels of income inequality (Chmielewski & Reardon 2016), which is rising in Norway. Moreover, achievement in Norwegian primary schools is strongly predictive of later educational success and ultimately life chances; among children achieving slightly below mean grade points by the end of primary school (i.e., < 34 points, 40 is the average), less than half completed secondary education, while among the top

[1] With contributions from Ane Nærde (Norwegian Center for Child Behavioral Development) & Mari V. Wang (Norwegian Institute of Public Health) who have kindly provided access to BONDS and MoBa datasets, respectively.

achievers (>50 points) 97% completed secondary education (Norway 2014). There are also significant differences in achievement scores in Norway attributable to school- and municipality-level factors beyond children's own family socio-economic conditions (Steffensen, Ekren, Zachrisen, & Kirkebøen 2017).

4.2 The present study

Within this context of a progressive, wealthy Northern European welfare state, albeit with increasing inequality and poverty rates, we investigate both the research questions outlined in the introduction chapter of this report. *First*, we ask when social and migration disparities in educational achievements arise and how they evolve over the early life stages. *Second*, we ask whether social and migration disparities in the early years fully explain differences at later points of the educational career, or whether these later differences are partly shaped by the role that SES and migration background may play beyond the early years.

We address these questions using two Norwegian datasets: the Behavior Outlook Norwegian Developmental Study (BONDS: Nærde, Janson, & Ogden 2014), and the Norwegian Mother and Child Cohort Study (MoBa: Per Magnus et al. 2016; Magnus et al. 2006). While having very different sample sizes (BONDS approximately n= 1,150; MoBa approximately n=100000), these studies are both drawn from the general population in Norway, follow children and their families across the earliest years through school start, and have used roughly the same measures. While BONDS's strengths are the relatively high participation rates and the low attrition, MoBa has the advantage of being a very large-scale study. However, both studies have population-based sampling with incomplete participation and are not entirely representative of the Norwegian population (BONDS is limited to five municipalities in southeast Norway, whereas MoBa is nationwide). The rate of invited families not participating in the study is 40% for BONDS and 60% for MoBa. Moreover, while both studies use similar, well-recognised measures of language problems, they have no available data on actual family income. Thus, when measuring household income, we rely on an indicator of economic hardship, available for BONDS, and a crude measure of self-reported income during pregnancy, available in MoBa.

The analyses by household income are then complemented using another critical indicator of SES, that is parental education. Note that the strategy for the measurement of parental education in this chapter deviates from the overall strategy outlined in the introduction of the report. More precisely, we rely on maternal education as a continuous indicator rather than a categorical measure of the combined parental education level (details in 4.3.2.1). Moreover, we include two indicators of migrant background in the BONDS analyses; whether parents have roots in a western- or non-western country apart from Norway. In MoBa, we use a single indicator on whether another language than Norwegian is spoken at home. While admittedly crude measures, they are in part used due to the diversity of the non-Norwegian population, having background from many countries, and with no immigrant group being a clear majority. This diversity has also had implications for ethical considerations, as none of the datasets has been allowed to ask specifically about the country of origin. Finally, as both datasets primarily rely on parent-reported information on young children, the measures are primarily addressing communication/ language problems (with the exception of BONDS having teacher ratings of school achievement in 1st and 2nd grade). We are thus constrained by both studies having early this dimension of cognitive development as their foci.

Notwithstanding the limitations, the two datasets are invaluable sources of information for examining social and migration differences in the early language skill development in Norway. What is more, the availability of two separate datasets compensate for the lack of representativeness of the data and allow us to be more confident in drawing policy implications if the results from the two datasets converge.

4.3 Methods

4.3.1 Participants

4.3.1.1 BONDS

The study includes three cohorts of children and their families (*N* = 1157) from five municipalities in southeast Norway. Using diverse data collection methods, data on cognitive, social, and behavioural development was gathered from 6 months to 2nd grade. Recruitment took place in three waves – in 2006 (*n* = 433), 2007 (*n* = 529) and 2008 (*n* = 195) – through public child health clinics attended by almost all families in Norway. Parents of 1931 eligible children, with at least one Norwegian-speaking parent, were informed about the study by a staff nurse, 1465 (76%) agreed to be contacted, and, subsequently, 1159 (60%) agreed to participate (two families later withdrew from the study and their data files were discarded, in a total of 1157). Retention rates were as follows: 12 months (98%), 24 months (95%), 36 months (93%), 48 months (93%), 1st grade (82%) and 2nd grade (78%), yet not all participants attended all parts of the data collection.

Teacher reports were obtained for 76% of the original sample in 1st grade, and for 77% in 2nd grade. For the purpose of the current report, we used information collected *via* parent interviews and questionnaires on communication and language development/difficulties (at 6, 12, 24, and 36 months), child expressive language testing (at 48 months), and teacher reports on language and overall achievement (1st and 2nd grade). The BONDS is approved by the Norwegian Social Science Data Services and the Regional Committee for Medical and Health Research Ethics (Protocol number: 2009/224. Study name: "BONDS").

4.3.1.2 MoBa

The population-based Norwegian Mother and Child Cohort Study (MoBa; for a complete description, see Magnus et al., 2006, and www.fhi.no/morogbarn) is a longitudinal, multi-cohort, health study in Norway. All pregnant women in Norway who received routine exams at birth units delivering more than 100 births per year were invited to participate during the 17th gestational week visit between the years of 2000 and 2010. In total, 95,369 mothers of 103,219 children had enrolled and completed baseline assessments, which represented 42.1 % of all eligible mothers in Norway. Written informed consent was obtained and the study was approved by The Regional Committee for Medical Research Ethics and the Norwegian Data Inspectorate.

Questionnaires covering demographics, health, lifestyle, and child development were administered during the 17th, 22nd and 30th weeks of gestation, and at ages 0.5, 1.5 and 3 years (questionnaires are available online, www.fhi.no/moba-en). The retention rate at 1.5 and 3 years was 72.4% and 59.3%, respectively, and 41% and 42%, at age 5 and 8 years, respectively. We included data from maternal reports during pregnancy (demographic data), and on language problems at child ages 6, 18, and 36 months, and at 5 and 8 years.

4.3.2 Measures

4.3.2.1 Demographic and family characteristics

BONDS

At age 6 months, parents were interviewed and asked about the child's gender, date of birth, immigrant status (recoded into *western* and *non-western*, used as dummy variables with Norwegian as the reference group), and maternal education (self-reported years of education). Note that the use of maternal education in the BONDS analyses deviates from the overall strategy used in other chapters. This is because the group of children having both parents with less than completed high-school education was very small (3.8%), thus rendering unstable results throughout analyses. By maintaining education as a continuous indicator, and relying on maternal education, rather than combined parental education, we balance statistical power and relevance. Economic hardship was assessed at age 12 months with a single indicator, where parents reported whether they had financial problems during the last 12 months (paying rent, loans etc.), responding "Yes" (1) or "No" (0).

MoBa

Demographic characteristics were reported by the mother during pregnancy. These include maternal education (the same question as used in BONDS), language background (used as a proxy for immigrant background in this report), and family income. Family income was reported in crude categories as pre-tax income. Within each cohort, we defined the lowest quartile as being "low income" and use this dummy variable in the analyses. Child gender was retrieved from the medical birth registry. Table A1 in the Appendix (Section 4.1) provides descriptive statistics for both the BONDS and MoBa data.

4.3.2.2 Dependent variables

Across ages, we use age-appropriate screenings for language problems, and at a one-time point in BONDS (48 months), a test of language skills. In addition, for BONDS in 1st and 2nd grade, we use teacher ratings of school achievement. These measures provide indicators for our latent dependent variables, outlined below. Table 1 provides an overview of measures at each time point.

Ages and Stages Questionnaire (ASQ)

We used the Norwegian version of the Ages and Stages Questionnaire, a series of child development screening questionnaires each made up of 30 items in five domains: Communication, Gross Motor, Fine Motor, Problem Solving, and Personal-Social. We used the social communication subscale of the Norwegian validated version of the ASQ (Ritcher & Janson 2007). To tap developmental risk appropriately, there are different versions of the scale according to the age of the child. An example of an item used when the child was six months old is: "Does the child make noises like "da", "ga", "ka", and "ba"?". At 24 months, a sample item is: "Without you first showing, does the child *point* at the correct picture when you say "Show me the kitty" or ask "Where is the dog"?" The child does only have to point at one correct picture. The items have three response categories: "Yes, very often", "Yes, sometimes", and "Not yet". The ASQ shows good test-retest agreement and concurrent validity (Squires et al. 1997). For the MoBa at 36 months, the items included four original 36-month ASQ items, and one item each from the 18- and 48-month ASQ questionnaires.

Children's Communication Checklist – 2nd rev. (CCC2)

The CCC2 was initially designed to be completed by parents, as a report on aspects of their children's communicative strengths and weaknesses that are not amenable to more conventional assessment (Bishop 2003). The original version contains 70 items divided into 10 scales. For the BONDS study, we included 28 items completed by teachers in 1st and 2nd grade. The 28 items were selected from the first four scales, assessing structural aspects of language: speech, syntax, semantics, and coherence. For each of these scales, five items (in total 20 items) describe weaknesses and two (in total 8 items) describe strengths. For the MoBa, a selection of 6 items was used in the 5year questionnaire, while 16 items were used at 8 years (including the same items used at 5 years). The data from age 5 was available for cohorts from 2004 through 2009, at age 8 from 2002 through 2009. An example of a "weakness" item on the "coherence" scale is: "It is hard to make sense of what she is saying (even though the words are clearly spoken)," and a "strength" item on the speech scale is "Speaks fluently and clearly, pronounces all speech sounds clearly and without hesitation". The teacher is asked to rate the frequency of each behavior, with the following response categories "Less than once a week (or never)", "At least once a week, but not every day", "Once or twice a day", or "Several times (more than twice) a day (or always)". We used items from the official Norwegian translation of CCC2 (Helland & Heimann 2007; Helland & Møllerhaug 2002), which was used under licensing agreement with the publisher Pearson.

Grammar scale (BONDS only)

We used a grammar scale developed by Dale and colleagues (2003), whose development was informed by the MacArthur Communicative Development Inventory: UK Short Form (MCDI–UKSF; Dionne, Dale, Boivin, & Plomin 2003). The measure was included in the Twins Early Development Study parent report booklet and has been able to identify children at risk for language impairment in several studies (Bishop, Laws, Adams, & Norbury 2006). We had data on this measure for ages at 12, 24 and 36 months, rated by an external trained coder based on 15 minutes video-recorded parent-child interaction. At 24 and 36 months, we also had ratings by the parents. According to the procedure, raters were asked to choose among sentences which best characterize the language skills of their children. For example, for 2-year-olds the sentences are: 1) not yet talking; 2) he/she is talking, but you cannot understand him/her; 3) talking in one-word utterances, such as "milk" or "down"; 4) talking in 2- to 3-word phrases, such as "give doll" or "me got ball"; 5) talking in fairly complete sentences, such as "I got a doll" or "can I go outside?"; 6) talking in long and complicated sentences, such as "when I went to the park, I went on the swings" or "I saw a man standing in the corner".

British Picture Vocabulary Scale–II (BONDS only)

Around the time children completed 48 months of age, children's receptive vocabulary was assessed with the Norwegian version of the British Picture Vocabulary Scale–II (BPVS–II; Dunn, Whetton & Burley 1997). The Norwegian version of this scale includes 12 of the 14 original sets of pictures. These main picture sets are ordered in ascending difficulty level and each includes 12 subtasks. The child is requested to choose the picture (out of 4) which best illustrates a word/concept uttered by the person administering the scale. The child can either point to the correct picture or indicate the picture number. Target words cover a wide range of language levels, as well as word classes, and are allocated to different semantic and/or grammatical groupings (actions, adjectives, animals, emotions, food and so on, cf. Dunn et al. 1997). The sum of all correct answers forms a raw score. For the current study, we used raw scores as opposed to standardised scores since no adequate norms have been developed for Norwegian samples. However, in the current analysis, we adjust for testing age, so the scores are normed to our sample (for further details about scoring procedures, see Zambrana, Dearing, Nærde, & Zachrisson 2015). Cronbach's alpha for the BPVS–II in our sample was .81.

Teacher-rated academic skills (BONDS only)

Near the end of the Fall semester in 1st and 2nd grade, children were assessed with the Social Skills Improvement System Rating Scales (SSIS–RS; Gresham & Elliott 2008). The SSIS–RS is designed to assess *social skills*, *problem behaviours*, and *academic competence* in children/youth from 3 to 18 years. This is a multi-rater series of rating scales that include ratings from teachers, parents, and students. In the current study, the subscale *academic competence* derived from teachers' ratings was used, which include areas such as reading/writing, math, motivation, and overall achievement. Teachers were asked how a given child rated in terms of expectations for their grade level, for example in reading/writing, when compared to their classmates, on a 5-point scale (1 = the lowest 10% of class, 2 = the next lowest 20%, 3 = the middle 40%, 4 = the next highest 20%, 5 = the highest 10%). High stability has previously been found in how teachers place children in these broad performance categories (Gresham & Elliott 2008). The validity of teacher-rated performance can be inferred from a meta-analysis of 73 studies, which found an overall correlation of .63 between teachers' ratings of student performance and standardised test scores (Südkamp, Kaiser, & Möller 2012).

Table 1 Measurements across ages in BONDS and MoBa.

			BONDS			
Age/grade	ASQ com.	Grammar Scale R1	Grammar Scale R2	BPVS–II	Teacher's ratings	CCC2
6 m	**X**					
12 m	**X**	**X**				
24 m	**X**	**X**	**X**			
36 m	**X**	**X**	**X**			
48 m				**X**		
1st g					**X**	**X**
2nd g					**X**	**X**
			MoBa			
6 m	**X**					
18 m	**X**					
36 m	**X**					
5 y						**X**
8 y						**X**

Notes: m = months, y = years, g = grade.

4.3.3 Analytical strategy

For both RQ1 and RQ2, we fitted measurement models of all dependent variables in both datasets (i.e., for BONDS, this includes constructs for communication/language problems/language skills, and school achievement at each time point, across ages 6 months to 2nd grade) by using SEM in *Mplus*. For BONDS, we had a complex set of measures. At age 6 months, we constructed a latent variable "Lang 6" including 6 items of the ASQ communication subscale. For ages 12 (Lang12), 24 (Lang 24), and 36 (Lang 36), items from the ASQ were also included in our latent variables in addition to the item(s) from the Grammar Scale (for one or both raters as available). For age 48 months, we used the raw score for the BPVS–II. For 1st grade, we fitted models measuring two separate constructs. In the first model, we constructed a two-factor model based on the teacher ratings of CCC2 in 1st and 2nd grade, respectively. The first CCC2 (LA) factor includes all negatively phrased items from the four subscales (speech, syntax, semantics, and coherence), the second CCC2 factor (LB) includes all the positively phrased items form the same subscales.

The second model included these two first-order factors (LA and LB) plus a third first-order factor *Teacher's Overall Ratings* (TOR), based on the ratings from SSIS–RS, and a second-order factor "Achievement" (ACH1) composed of the three first-order factors. For 2nd grade, the same procedure was carried out with one model with two factors (LA and LB) plus a second model with three first-order factors (LA, LB and TOR2) and a second-order factor (ACH2). For the MoBa, we used ASQ items only at ages 6, 18, and 36 months, and CCC2-items at ages 5 and 8 years. For both datasets, we addressed RQ1 by examining at which ages social and migration gaps arise in a cross-sequential manner. More precisely, we regressed the latent language/achievement score on indicators of gender, deviations from expected age of testing at each time point (for BONDS only), maternal education, economic hardship, and non-western immigrant background, as well as design-specific sites (BONDS only) and birth cohort dummies. Missing data were handled through standard missing estimation procedure in *Mplus* (full information maximum likelihood estimation when ML estimator was used, and equivalent procedure when WLSMV estimator was used). To ease interpretability of our tables and figures, we inverted the predictions for the two positively loaded dependent variables (BPVS–II at age 48 months in BONDS and the second CCC2-factor in both datasets), so that a higher score on the dependent variable always reflects more problems.

For RQ2, we analysed the role of early inequalities for the emergence of inequalities in the 1st and 2nd grade of primary school in BONDS, and age 5 and 8 in MoBa, respectively. We fitted an overall latent factor (second order) including all communication/language measures prior to school age, to capture overall early skills. For the latest measures, we estimated a latent change model for language delay, including the two latent variables from CCC2 (in 1st and 2nd grade for BONDS, 5 and 8 years for MoBa), estimating intercept and change, after first establishing strong factor invariance across these two time points to ensure that we measured true change. We then regressed the intercept and change score on the early skills measure, indicators of maternal education, economic hardship, non-western immigrant status, and site (BONDS only) and cohort.

4.4 Results

4.4.1 The development of social and migration gaps in achievement over the early life course (RQ1)

Table 2 shows details for all measurement models across age, including the number of items, the range of factor loadings, and the model-based fit indexes. All our measurement models fitted the data well, and there were no correlated errors.

Table 2 Fit indexes for measurement models for both datasets.

	Number of items	Range of factor loadings	CFI	TLI	RMSEA
BONDS					
Lang 6 m	6	.20–.39	.97	.94	.03
Lang 12 m	6	.37–.58	.97	.95	.04
Lang 24 m	6	.41–.84	.97	.94	.08
Lang 36 m	6	.47–.83	.99	.98	.04
Two-factor model (1st grade)					
LD1	19[a]	.60–.90	.93	.92	.08
LS21	8	.73–.96			
Two-factor model (2nd grade)			.95	.95	.06
LD2	19[a]	.70–.95			
LS2	8	.77–.93			
Hierarchical model (1st grade)			.92	.91	.07
LD1	19[a]	.61–.91			
LS1	8	.74–.95			
TOR1	4	.84–.95			
ACH1	–	.65–.88			
Hierachical model (2nd grade)			.95	.94	.05
LD1	19[a]	.77–.94			
LS1	8	.77–.93			
TOR1	4	.80–.93			
ACH1	–	.64–.90			
MoBa					
Lang 6 m	5	.33–.72	.96	.91	.02
Lang 18 m	3	.62–.89	1.00	1.00	.00
Lang 36 m	6	.7–.89	.99	.99	.03
CCC2 5 y	5[b]	.31–.75	.99	.98	.04
CCC2 Two-factor model (8 y)					
LA	9	.60–.83	.95	.94	.06
LB	7	.58–.92			

Notes: [a] The item # 20 was discarded due to convergence problems, [b] Fit index after removing one of the two "strength" items with a weak (>.20) factor loading. Due to only 5 items, only an LA factor was tested for CCC2 at age 5.

Starting from our measurement models as a baseline, we included predictors for gender, exact age at testing (BONDS only), maternal education, economic hardship and migration background. Results are presented in Table 3 and Table 4. Looking at the results for BONDS (Table 3), boys seem to lag behind in the development of language and exhibit more language problems than girls, consistently from 12 months of age to the 2nd grade of primary school (the only exception is the better performance of boys in expressive language at 48 months). Results are quite similar for MoBa data (Table 3): boys seem to have more language problems than girls all along the early life course. As expected, the age of testing (BONDS only) is negatively and significantly associated with language skills throughout the observation window (older children perform, on average, better). Importantly, these initial results indicate that our measurement models reflected our theoretical constructs with meaningful criterion validity.

Results regarding the role of maternal education and economic hardship are interesting and somewhat surprising. In the first two observation points of BONDS (6 and 12 months) and the first time point of MoBa (6 months), higher maternal education seems to be associated with more communication problems. The same pattern is apparent with regard to economic hardship, at least when looking at the results from the BONDS data (both at 6 and 12 months of age): children whose parents report more economic hardship seem to experience fewer communication problems. This latter result does not hold for MoBa, however. Given the striking finding, we tested the models excluding stepwise each of the predictors to avoid suppressor effects, of which we found no evidence. We also tested for measurement invariance between children of mothers with the lowest levels of education (not completed high-school) and the rest. Measurement invariance test suggests that this striking finding is not attributable to measurement artefacts. Regarding migration background, we find no association with communication problems during the first year of life (6 and 12 months), at least looking at BONDS data. However, results from MoBa suggest that children with a migration background experience more language problems all along their early life courses.

The surprising positive association between SES and language problems vanish when looking at the same children after 24 months after birth (BONDS). While there is no clear evidence that SES (maternal education and economic hardship) or migrant background is associated with language problems, the point estimates suggest that 2-year-olds with higher maternal education may experience fewer communication problems ($p < .1$). The results from MoBa, which are based on a significantly larger sample size, confirm this reversed trend and show a negative association between maternal education and language problems from 18 months onwards, albeit weak. For example, after 18 months, 1 SD increase in maternal education (about 2.6 years) is associated with only a .04 SD decrease in language problems.

In both datasets, the negative association between maternal education and language problems gets stronger at 36 months of age, and remains persistent throughout the observation windows, until 2nd grade in the case of BONDS and 8 years of age in the case of MoBa (the only exception is the second CCC2 factor in first grade in BONDS, where we found no significant association). Standardised regression coefficients range from –.04 to –.14 in both studies, meaning that a one SD increase in maternal education (about 2.5 years) is a 4% to 14% SD reduction in communication-related problems. Notwithstanding the stability of the negative sign of the relationship, the effect sizes vary across ages without any consistent pattern in either of the studies. Worth noting is that the size of the coefficients over time cannot be formally compared since the measurement of the outcome different across time points. Thus, we cannot establish with certainty whether the actual gaps in language problems fluctuate over the early life course or whether this fluctuation is a function of the measures/items used from time to time. It is, however, worth noting that the strongest associations with regard the two measures capturing a wider range of language skills rather than merely language problems (BPVS–II in BONDS, and CCC2 LB in both studies). This latter result suggests that there are meaningful differences related to maternal education across the full range of language skills and that these differences are not restricted to language problems captured by the screening measures used at most time points.

Apart from the striking positive association after 6 and 12 months after birth (see above), economic hardship is not related to language problems in all other occasions before the school entry (BONDS). Yet, moving from preschool to the first 1st and 2nd grade in BONDS, and from age 36 months onwards in MoBa, children whose parents reported economic hardship during their first year of life perform consistently worse than their peers who had not had this experience (with the exception of the LB capturing language strengths in MoBa). When we convert standardized regression coefficients (included in Table 3) in Cohen's *d* (the group difference in SD units), we see that the difference between the performances of children who experienced economic hardship and those who did not ranges from 26% to 40% of a SD in the BONDS sample (26% and 40% for the two CCC2 factors in 1st grade, and 38% and 32% of a SD in 2nd grade, respectively). While we found no association between migration background and language problems during the first year of life in the BONDS sample, non-western immigrant children seem to have more language problems at 36 and 48 months of age compared to natives. However, the latter result holds only for one out of the two CCC2-factors (LA, the problems factor) in 1st grade and none of the factors in 2nd grade. But again, results from MoBa show that children with immigrant background are consistently rated as having more language problems than their native peers over the early life stages.

Table 4 shows coefficients for the models analysing the overall achievement in 1st and 2nd grade using the BONDS sample. On average, boys score lower than girls and older children at the testing score higher than younger children on these overall measures of achievement. Maternal education is positively associated with the overall achievement, meaning that children of mothers with higher education score higher than children of low educated mothers. Yet, children of parents who did not report economic hardship during the child's first year of life score higher on the overall measure of achievement compared to their peers who experienced economic hardship.[2]

[2] Fit indexes for all models presented in Tables 3 and Table 4 are displayed in Table A2 in the Appendix 4.2. All models except fitted the data well, except the models at 6 and 36 months of age on the BONDS sample. Therefore, results from these two time points should be interpreted with caution.

Table 3 Predictors of communication problems / poor language development from 6 months to 2nd grade (BONDS) / 8 years (MoBa).

				BONDS					
Age/grade	6 m	12 m	24 m	36 m	48 m	1st g		2nd g	
N	1071	1077	1042	1032	897	840		860	
Outcome	Lang 6	Lang 12	Lang 24	Lang 36	Lang 48	LA	LB	LA	LB
Child gender	–0.02	0.15***	0.24***	0.37***	–0.01	0.22***	0.18***	0.18***	0.18***
(boy)	(0.05)	(0.04)	(0.03)	(0.07)	(0.03)	(0.04)	(0.35)	(0.04)	(0.04)
Exact age at	–	–0.25***	–0.14**	–0.16**	–0.08*	–0.15***	–0.15***	–0.14**	–0.15***
testing		(0.04)	(0.03)	(0.06)	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)
Maternal	0.15**	0.17***	–0.06^{+}	–0.05***	–0.14***	–0.11*	–0.04	–0.10*	–0.14***
education	(0.05)	(0.04)	(0.03)	(0.01)	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)
Economic	–0.13**	–0.08*	0.02	0.07	0.02	0.08*	0.13***	0.12**	0.10**
Hardship	(0.05)	(0.04)	(0.03)	(0.11)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Non–West	–0.06	0.03	0.06^{+}	0.66***	0.08*	0.08*	0.26	0.07	0.04
immigrants	(0.05)	(0.04)	(0.03)	(0.15)	(0.03)	(0.04)	(0.17)	(0.04)	(0.04)
Western	0.03	–0.01	0.07*	0.37**	0.05	–0.01	0.02	0.03	0.02
immigrant	(0.05)	(0.04)	(0.03)	(0.14)	(0.03)	(0.04)	(0.15)	(0.04)	(0.04)
				MoBa					
Age	6 m	18 m		36 m		5 y		8 y	
N	80,372	68,483		54,568		38,466		40,111	
Outcome	Lang 6	Lang 12		Lang 36		LA		LA	LB
Child gender	0.03***	0.24***		0.16***		0.14***		0.08***	0.12***
(boy)	(0.07)	(0.01)		(0.01)		(0.01)		(0.01)	(0.01)
Maternal	0.05***	–0.04***		–0.12***		–0.06***		–0.10***	–0.13***
education	(0.01)	(0.01)		(0.01)		(0.01)		(0.01)	(0.01)
Economic	0.01	–0.00		0.03***		0.03***		0.02***	–0.01
hardship	(0.01)	(0.01)		(0.01)		(0.01)		(0.01)	(0.01)
Immigrant	0.04***	0.04***		0.06***		0.08***		0.07***	0.05***
	(0.01)	(0.01)		(0.01)		(0.01)		(0.01)	(0.01)

Notes: Higher score on the dependent variable means more problems. All coefficients are standardized on the x and y variable. Standard errors in parentheses. Significance: $^{+}$ p<.1; *p< .05; **p < .01; ***p<.001.

Table 4 Predictors of overall achievement 1st and 2nd grade (BONDS).

	1st grade	2nd grade
Child gender (boy)	–0.23***	–0.21***
	(0.04)	(0.04)
Exact age at testing	0.18**	0.18***
	(0.04)	(0.04)
Maternal education	0.11**	0.15***
	(0.04)	(0.04)
Economic hardship	–0.12**	–0.15**
	(0.04)	(0.04)
Non–West immigrant	–0.10**	–0.06
	(0.04)	(0.04)
West immigrant status	0.00	0.00
	(0.04)	(0.04)

Notes: BONDS data only. All coefficients are standardized on the x and y variable. Higher score on the dependent variable means higher achievement. Standard errors in parentheses. Significance: *p< .05; **p < .01; ***p<.001.

4.4.2 Explaining social and migration gaps in primary school by preschool achievement (RQ2)

Our second research question investigates the role of early inequalities for later inequalities and whether social and migration background play a role in shaping achievement inequalities beyond the early years. In the BONDS data, early inequalities are measured by tests taken prior to school age while later inequalities refer to the first two years of schooling. In MoBa, early inequalities are measured by tests taken during the first three years of life while later inequalities are measured by tests taken from age 5 onwards (school entry happens the year children turn 6, so at age 5 some children are measured at the beginning of 1st grade, some just before school entry). Since the measures in the early years are restricted to language and communication problems, we included only our measure of language in later time points (CCC2; LA and LB) in the analysis of both datasets. We then control for roughly the same domain of skills as we include in our prediction model.

We expand the initial research question by estimating a score for the change in language problems from 1st to 2nd grade in BONDS, and from age 5 to 8 years in MoBa. Thus, we predict the initial level of language problems in 1st grade/age 5 from earlier measures of language problems, maternal education, economic hardship, and immigrant background (in addition to gender, age at testing, and site- and cohort dummies), as well as change in language problems from 1st to 2nd grade in BONDS and from age 5 to 8 years in MoBa, above and beyond the initial status in 1st grade predicted from the same set of independent variables.

4.4.2.1 Measurement and latent change models

Our initial step is the estimation of a model of intercept and change in the two subscales of CCC2 (LA and LB) measuring language problems in 1st and 2nd grade, which are our outcome measures in BONDS. Due to the selection of items at age 5 in MoBa, we restricted these analyses to include only the LA (problems) subscale. In order to ensure that we measure the same constructs over time, we tested measurement invariance. This procedure involved the estimation of free models with both time points, in which factor loadings and thresholds from the factor model for CCC2 (described above) are freely estimated, with the residuals of identical items correlated over time. In the case of BONDS, we needed to make some adjustments. More precisely, we removed three items with empty cells (i.e., items with no responses for specific values across time) that were responsible for the failure of model convergence. The items were #10, #17, and #24 (collapsing values with few observations was not successful). The model fitted the data well (see Table 5). When we constrained step-wise the factor loadings and then the thresholds, both models showed no reduction in model fit. Hence, we can assume that the construct is measured in the same way across the two time points, thus allowing us to estimate a real change. The model with constrained loadings and thresholds was retained for further analyses.

In MoBa, we estimated measurement invariance for the 4 items measured both at 5 and 8 years, while also including the remaining items measures at age 8. Unfortunately, we were unable to achieve strict measurement invariance for all items and had to let one item ("it is hard to make sense of what he/she is saying…") to vary freely over time. Moreover, we had to allow the first threshold (between ratings of 1 and 2) on two of the items to be freely estimated over time to reach invariance. Hence, the estimated stability and change in the MoBa may be contaminated with some measurement error (see Table 5).

In a second step, we estimated a latent model capturing language problems prior to school age. We fitted a hierarchical factor model (second-order factor model) where each of the ASQ factors described above (at 6, 12, 24, and 36 months in BONDS, 6, 18, and 36 months in MoBa) were included. In both datasets, the model fit was good. For example, in BONDS, Chisq [270] = 650.55, $p < .000$, RMSEA = .03, CFI = .92, TLI = 91, with second-order factor loadings of .21, .51, .88, and .83, respectively for Lang at 6, 12, 24, and 36 months (all $p < .000$).

Table 5 Fit indexes measurement invariance testing of i) a two-factor model of CCC2 across 1st and 2nd grade (BONDS); ii) a one-factor (problems) model of CCC2 across at ages 5 and 8 (MoBa).

	Df	Chi^2	RMSEA	CFI	TLI
BONDS					
Free model	1144	2214.59***	.03	.97	.97
Constrained factor loadings	1167	2103.94***	.03	.97	.97
Constrained thresholds	1242	2438.70***	.03	.97	.97
MoBa					
Free model	60	2272.27***	.03	.99	.98
Constrained factor loadings [a]	62	3953.31	.03	.97	.96
Constrained thresholds [b]	69	4126.97***	.03	.97	.97

Notes: [a] One item estimated freely. [b] The first threshold on two items estimated freely. Significance: *p< .05; **p < .01; ***p<.001.

4.4.2.2 Prediction models

The main estimation model (displayed as a path diagram) is shown in Figure 1 (BONDS) and Figure 2 (MoBa). The latent variable DiffLang is the change (difference score) in language problems in the two language constructs measured by CCC2 (LA and LB in BONDS, LA in MoBa). Although presented in two separate models for ease of interpretation, we estimated the two outcomes simultaneously in one structural model in BONDS. The over-all model fitted the data well in both BONDS (Chisq [3595] = 5198.03***, RMSEA=.02, CFI=.95, TLI=.95) and MoBa (Chisq (522) = 15932.79***, RMSEA=.02, CFI=.94, TLI=.94).

The model in both datasets should be interpreted as follows. The DiffLang-variable represents the change from the first to the second time point (1st to 2nd grade in BONDS, age 5 to 8 years in MoBa) in language problems and should be interpreted as a slope in a growth model. LA/LB 1st grade (1st gr)/5 years (5y) should be interpreted as the intercept, i.e. the initial level of language problems in 1st grade/at 5 years. The models suggest that about one-third of the variance at the second time point in both datasets is due to change, while two-thirds of the variance is stable from the first time point. The correlation (double-arrowed line) between the first time point and DiffLang is the correlation between the initial level and the change. The correlation between the initial level and change (i.e., difference) in both models indicates that children starting with more problems in 1st grade have higher rates of change (e.g., in BONDS, it means that about 34% and 31% of the variance [*r*=.59 for LA and *r*=.56 for LB], respectively, in the change to 2nd grade is accounted for by the child's initial level of problems).

On the prediction side of the model, we estimated associations between a) early language problems, b) maternal education, c) economic hardship, and c) non-western and western immigrant background in BONDS, immigrant background in MoBa (contrasted to Norwegian background), and the initial level of language problems in 1st grade (LA/LB) and change in language problems to 2nd grade (DiffLang). Substantively, the models estimate the extent to which these predictors are 1) associated with the initial level of language problems in 1st grade/at age 5 years (LA/LB), and 2) with a change in language problems to the second time point, conditioning on the initial association with 1st grade/age 5 years problems.

Prediction models in BONDS

Interestingly, despite their high correlation, a slightly different pattern is apparent for the two factors (LA and LB). Keeping in mind that LA was comprised of negatively phrased questions about the child's speech (syntax, semantics, and coherence), Figure 1 (upper panel) shows the following pattern of predictions. Early language problems are by far the strongest predictors of language problems in 1st grade (est. = .79, $p < .001$, i.e. 1 SD higher score on early language problems was associated with 79% of a SD higher score on problems in 1st grade). However, these early measures do not predict change in language problems from 1st to 2nd grade. In addition, maternal education has a separate additive influence on top of earlier language problems (est. = –.10, $p < .05$), This means that above and beyond earlier language problems, a 1 SD higher maternal education (approximately 2.5 years) is associated with 10% of a SD less language problems. However, also maternal education is unrelated to change in language problems from 1st to 2nd grade. Finally, non-western immigration status is positively associated to language problems in 1st grade (est. = .11, $p < .01$) but has no influence on the change from 1st to 2nd grade. None of the other predictors is significantly associated either to language problems in 1st grade or the change from 1st to 2nd grade.[1]

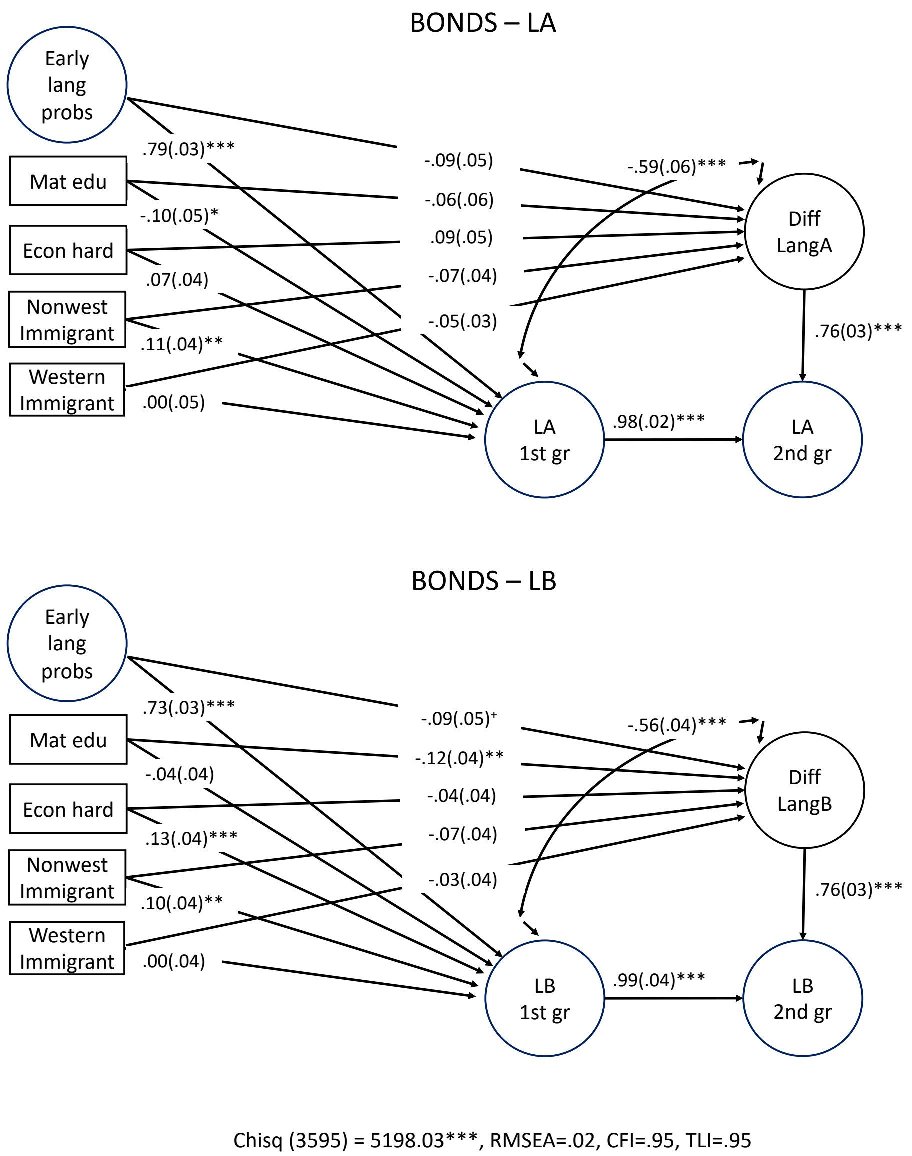


Figure 1 Latent change model showing associations between indicators of low SES (maternal education, economic hardship, and immigrant background) and changes in language skills from 1st to 2nd grade, conditioning on earlier language skills (BONDS).

Results for the second latent language score (LB, good language skills) in BONDS yield a somewhat different pattern. Worth noting is that, in Figure 1 (lower panel), coefficients of the prediction model are inverted, and therefore negative signs imply more language problems (the

[1] In a final step, we re-estimated the model excluding maternal education and economic hardship, respectively, to ensure that our results were not affected by multi-collinearity between the two measures of SES. The results are similar to those presented above, however. Moreover, as the two factors we identified in the initial measurement model were highly correlated (.73), we re-estimated two structural models including each of the two factors, respectively. Also, in the latter case, the results are identical.

coefficients in the latent change model are kept in their original direction as these should be interpreted as stability and change). Earlier language problems is the strongest predictor of problems in 1st grade: the coefficient is even similar (est. = .73, $p < .001$) to the one estimated for LA. However, while maternal education is not associated with language problems in 1st grade, both early economic hardship and non-western immigrant status are (est. = .13 and .10, $p < .001$ and $p < .01$, respectively). These coefficients convert to Cohen's *ds* of .43 for both variables. This means that children exposed to economic hardship or having immigrant background were rated as having 43% of a SD more problems than their more affluent, or Norwegian, counterparts, respectively. Importantly, and in contrast to the model measuring LA, maternal education was negatively associated with the 1st to 2nd grade change in language problems measured by LB. The standardised coefficient of –.12 ($p < .01$) implies that 1 SD higher maternal education (2.5 years) is associated with 12% of a SD reduction in the change.

Prediction models in MoBa

Keeping in mind that the MoBa data are restricted concerning numbers of items available from the CCC2, especially at age 5 years, and the slightly weaker invariance model, the coefficients presented in Figure 2 should be interpreted with caution. As in BONDS, LA stems negatively phrased questions. The main results are as follows. Early language problems is again by far the strongest predictor of language problems in 1st grade (est. = .52, $p < .001$, i.e., 1 SD higher score on early language problems is associated with 52% of a SD higher score on problems in 1st grade). Note that while this is a weaker prediction compared to BONDS, the measure at 5 years of age is also weaker.

In contrast to the findings from BONDS, early language problems also predict a positive change in language problems moving from age 5 to age 8 (i.e., more growth in problems). Maternal education is partially associated (controlling for earlier language problems) to both the initial level of problems and change. On the one hand, the coefficient of –.07 ($p < .05$) related to the initial level of problems implies that, above and beyond what is accounted for by earlier language problems, a 1 SD higher maternal education (approximately 2.5 years) is associated with 7% of a SD less language problems. On the other hand, the coefficient related to change (–.07, $p < .05$) implies that children of higher educated mothers have less increase in problems from age 5 to 8. Finally, economic hardship (i.e., low family income) and migration background are both associated with more language problems at age 5 but do not predict change in problems moving from age 5 to age 8.

In sum, maternal education was related to language problems at age 5 (BONDS) and in 1st grade (MoBa), above and beyond early language problems, and also associated with change in language skills to age 8 (MoBa) and 2nd grade (LB, inverted in Figure 1) for one of the outcomes (BONDS). Economic hardship was related to language problems in one of the models (LB) in BONDS, as well as in MoBa, while only associated with the change in problems in MoBa. Finally, non-western migrant status was associated with more language problems in both datasets, but not to change in problems, in any of the models.

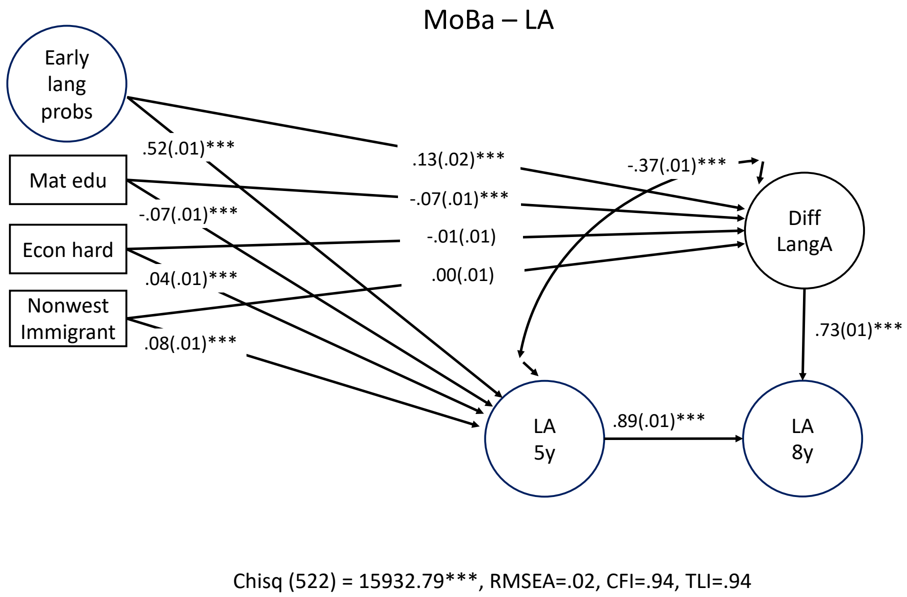


Figure 2 Latent change model showing associations between indicators of low SES (maternal education, low family income, and immigrant background) and changes in language skills from age 5 to 8 years, conditioning on earlier language skills (MoBa).

4.5 Conclusions

Relying on two Norwegian datasets, we investigated when social and migration gaps in educational achievements arise, how they evolve over the life cycle, and the role of early inequalities in shaping later disparities in achievements. We found that maternal education, economic hardship and migration background are associated in the expected direction with language/communication problems from ages 2–3 years through 2nd grade in BONDS, and from 18 months in MoBa. More specifically, in BONDS, the negative association between maternal education and children's language problems emerge around age 3 and seems persistent through 2nd grade. In MoBa, the same pattern was apparent starting from 18 months of age through age 8. However, In BONDS, the detrimental effect of economic hardship begins to show only from 1st grade onward, whereas the negative association between migration background and language/communication problems starts from age 2 and seems to wear off in 2nd grade, possibly reflecting the ability of the school system to compensate for family ethnic/language background. In MoBa, the association between language problems and low family income is apparent from age 36 months, and the negative influence of migrant background on language development is evident all along the early life course of children.

While starting from age 18 months onward the association between maternal education and language problems is in the expected direction, the same does not for children aged 6 and 12 months in BONDS, and 6 months in MoBa. At those life stages, higher maternal education is associated with more communication problems. Moreover, also economic hardship seems to be negatively associated with language problems prior to school age in BONDS and 36 months of age in MoBa, respectively. These latter findings are puzzling and require a few alternative explanations.

The first alternative explanation concerns measurement. As we tested and confirmed measurement invariance between mothers with the lowest levels of education (less than completed high school) and those with more education at both time points in BONDS, within-age differences in the measures do not seem to be the explanation. However, despite good model fit, low factor loadings in the language problems measures at ages 6 and 12 months suggests that the constructs at these time points are less precisely measured than at later ages, as there is more random noise in the measures. It is not clear how this affects the measures, just that the variance of the indicators is to a lesser extent than at later ages causally related to the constructs of interest (language problems). It is implausible that this affects the direction of the results merely than the standard error and, at most, the strength of the association. It is of course possible that the items of the ASQ at these early stages capture development not affected by SES. Yet, while the 6-month items address very early social communication (and to some extent senso-motoric precursors of social communication), the same is not the case for the items at 12 months (used in BONDS), which are strongly related to early language use (at age 12 months in BONDS, we also have one item externally rated by a trained observer). The validity of the measures is therefore not a plausible explanation either. An alternative hypothesis, relating to both datasets, may be subject to selective inclusion of parents, given the baseline participation rate of about 60% in BONDS and 40% in MoBA, and a skewed sample based on parental educational attainment (P. Magnus et al. 2006; Nærde et al. 2014). It is, therefore, possible that the most resourceful of the low educated parents were the ones committing to participate in the study. However, this is an explanation we cannot thoroughly test.

Regardless of the surprising results concerning the first year of life, the pattern reversed from somewhere between age 18 and 36 months and onwards and are generally consistent with previous findings reviewed elsewhere in this chapter. Moreover, while comparisons of coefficients based on different measures are inherently challenging to interpret meaningfully, the confidence intervals (i.e., +/– SE*1.96; not shown in the tables or figure) for the coefficients across time after age 36 month are in most cases overlapping. Thus, assuming comparable latent measures over time, there is no evidence for either widening nor narrowing gaps between age 36 months and 2nd grade/8 years.

This observation (though admittedly a contentious one, assuming comparable measures), draws attention to the analyses of RQ2 investigating the role of early inequalities for later inequalities and the role of social and migration background beyond the early years. Our findings suggest that, while early communication/language problems are strongly predictive of later problems, maternal education (with regard to the problem-focused latent CCC2-variable LA) makes an additional contribution to explaining problems in 1st grade (BONDS)/age 5 (MoBa). It is also notable that beyond the initial prediction of problems in 1st grade, none of the predictors explained any of the change in problems through 2nd grade in BONDS, while this was the case in MoBa. In contrast, for the other latent CCC2-variable (LB) measured in BONDS, which was an inverse of positively phrased questions, thus capturing resources in its original metric, maternal education was significantly related to change in problems through the first years of schooling, but not to initial level of problems in 1st grade. Interestingly, for this outcome, both early economic hardship and non-western immigrant background (positively) predicted problems in first grade, but not change in problems to 2nd grade in either of the datasets.

In conclusion, our analyses of the BONDS and MoBa data suggest that, in Norway, children of socioeconomically less advantaged parents start life with a (puzzling) advantage in terms of having fewer communication problems, while this reverses somewhere between ages 18 and 36 months. Through at least 2nd grade, a fairly stable socio-economic gradient remains. Moreover, early communication/language problems is a strong predictor of problems reported by the teacher after school entry, but not for the change in such problems throughout 2nd grade in BONDS, while it is predictive of change across a longer time span in MoBa.

5 UNITED KINGDOM

The Evolution of Social and Ethnic Inequalities in Cognitive Achievement from Preschool to Secondary Schooling in the UK

Jan Skopek♠ & Giampiero Passaretta♠
♠Trinity College Dublin

5.1 Introduction

This chapter will analyse social and ethnic achievement gaps for the United Kingdom. In the context of the overall report, UK represents an interesting contrasting institutional context to the more conservative welfare states of Germany and Italy and the universalistic context of Norway. In terms of the educational system, UK contrasts with its comprehensive schooling approach the tracking systems in Germany, Netherlands, and Italy.

Drawing upon longitudinal data from the Millennium Cohort Study, the chapter will address both questions raised in the report's introduction. First, we ask how socio-economic and migration-related inequalities in cognitive achievement unfold and evolve over the early years of life until the end of lower secondary education. Since migration and ethnic background are increasingly decoupling as UK is becoming an increasingly diverse society, our analysis distinguishes between children's migration background and children's ethnicity as two related but also, to some extent, separate dimensions. More precisely, we will provide answers to the following questions of (1) *when* in the early life course do socio-economic, migration background, and ethnicity gaps in cognitive achievement emerge in children; (2) *how large* are those gaps when children enter school age; (3) how do these gaps *develop* during school age? Moreover, our analysis also aspires to inspect some of the mechanisms that drive differential achievement of migrant and non-migrant children and children of different ethnic background.

Second, we will study, to which extent achievement inequality observed later in children's school life is a result of achievement inequality in the earlier years before children entered school. More precisely, we will make use of rich longitudinal data to estimate how much of early years achievement inequality is carried over to achievement inequality in school age. By that, we aim to disentangle the effects of socio-economic, migration- and ethnicity-related background characteristics on school achievement into a part that operates *before* and an additional part that operates *after children enter school*. Such a disentanglement bears important implications for social policies aiming to reduce achievement inequalities in school.

We will investigate the questions by looking at different outcomes including verbal skills, quantitative skills, and a composite measure of various abilities and skills. Moreover, achievement in school will be analysed by exploiting data from additional evaluations that were done by children's teachers. The analyses in this chapter build upon previous research which has analysed the MCS in terms of inequality by parental education (Bradbury, Corak, Waldfogel, & Washbrook 2015) or migration and ethnic background (Hoffmann 2018). Hence, the chapter inevitably replicates some of those earlier findings which is vital for the sake of the overall comparative approach of the report. Nonetheless, it contributes by combining earlier and adding new partly more nuanced perspectives on the evolution of achievement gaps in the UK.

The chapter is structured as follows. First, we will give a brief overview to the institutional setup of the education system in the UK. We will also briefly discuss the specific ethnic composition of minority groups in the United Kingdom. Afterwards, we introduce the data from the Millennium Cohort study and our analytical strategy. The results part will present findings on socio-economic as well as ethnic and migration-related achievement gaps in the UK. The chapter closes with a short conclusion.

5.2 The education system in the UK

The education system in the UK is split into early years education, primary schooling, secondary schooling, further education and higher education. Compulsory schooling starts at age 5 (in Northern Ireland at age 4) and lasts till age 16 by which the General Certificate of Secondary Education exam (GCSE) is taken (at least in England, Wales and Northern Ireland; in Scotland national standard grades). Like other countries in Europe, such as the federally organised systems of Germany or Switzerland, education in UK is decentralised on multiple levels. The most important level of decentralisation in the UK is the constituent country.

5.2.1 Early years education

British early formal childcare in the age groups below three is characterized by low affordability, low availability, and low utilization, mostly as a result of high childcare costs (European-Commission/EACEA/Eurydice/Eurostat 2014; OECD 2016c, 2017; West 2006). In European wide comparison, ECEC participation under age of three is low in both terms of overall rates and intensity of utilization (in 2011, about 30% of children visited formal childcare for less than 30 hours a week and just 5% more than 30 hours). Childcare provision for children younger than three heavily relies on market-based services with a rather low quality. As a result, there is considerable inequality in utilization of early childhood education and care services below age three: children from lower income families (although single-parents enjoy subsidies), and particularly children from immigrants, have substantially lower attendance rates at ECEC services. However, ECEC participation is high (more than 95%) for the preschool years between age 3 and age 5, possibly due to the charge-free provision of early education for children of aged 3 and 4 (which has been introduced stepwise since 2000) provided in England and, likewise, in Scotland and Wales. Note that, for the observation window of our study (children born in the UK from 2000 to 2001), the entitlement to free early education was limited to part-time education for four-year old children only. By age of 5, the starting age of compulsory schooling, nearly 100% of children are enrolled to primary education.

5.2.2 Primary and secondary schooling

The education system in the UK is decentralized and features some important heterogeneities across the constituent countries (for an in-depth discussion see West et al. 2010). The UK government oversees education only in England, while education is a local matter in Scotland, Wales, and Northern Ireland. England and Wales are rather similar in terms of education systems. Instead, large institutional differences can be found between England – with a quasi-market of schools featuring ‘league tables’, a partially selective school setup, an emphasis on school autonomy and parents’ free school choice – on the one hand, and Scotland – featuring less school autonomy and between school competition, as well as a fully comprehensive (i.e., not selective) approach (West et al. 2010), on the other. Since England is by far the largest country in the UK (about 84% of the overall population followed by Scotland with 8%), our discussion puts focus mainly on England's education system while acknowledging some differences to the other countries.

Compulsory schooling starts during the school year when children turn five and is organised into four so called Key Stages. Key Stage 1 (age 5 to 7) and Key Stage 2 (7 to 11) mark primary school education. Key Stages 3 (11 to 14) and 4 (14 to 16) represent secondary schooling, which terminates with the GCSE exams in the last year. After that period, students can progress into post-16 education (such as A-levels), which is typically preparing for later college education.

During the 1960s and 1970s the UK abandoned its highly tracked education system in favour of a comprehensive schooling approach (Manning & Pischke 2006). Although nowadays education in England no longer involves formal tracking, as found for example in Germany or the Netherlands, there are some significant differences between schools in terms of funding and specialisation. Approximately 93% of all students between age 3 and 18 visit state-funded schools, which are free of charge and mostly organised by local authorities. A small fraction of students (7%) attend privately-organised and fee-charging independent schools which have a selective student intake and do not follow the National Curriculum standards. Most of secondary schooling is comprehensive in the sense that schools do not select on students' abilities. However, even among the state-funded schools, grammar schools (about 5% of secondary schools) select students based on their academic abilities (based on results of a grammar school entrance exam at end of Key Stage 2). In contrast to the English system (and the system Northern Ireland), which retained selective elements despite the system transformation in the 70s, secondary schools in Scotland and Wales are fully comprehensive, i.e. non-selective.

Next to selective secondary schools, educational differentiation manifests through curriculum (subject) choice and school specialisation. Previous research on subject choice in English secondary schools revealed considerable gender, ethnic and social disparities in curriculum choice (Jin, Muriel, & Sibieta 2010; Sullivan, Zimdars, & Heath 2010). Parental education was found to be a strong predictor of students' choices of academic rather than vocational subjects even after taking into account students' ability levels, thus being consequential for later achievement and educational decisions (McMullin & Kulic 2016). Social disparities in curriculum choice are consequential for later labour market outcomes too – for example, differences in the curriculum choice explain, to a large extent, social class differentials in occupational attainment (Iannelli 2013). Similarly, there is a strong association between social origin and subject choice in Scotland, with upper class students taking more frequently subjects that facilitate access to higher education (Klein, Iannelli, & Smyth 2016).

5.3 Data

5.3.1 The Millennium Cohort Study

We used data from the Millennium Cohort Study (MCS). The MCS is the first nationally representative birth cohort study on children born in the UK in 2000 to 2001 (Hansen 2014). It covers children from the four UK countries England, Northern Ireland, Scotland, and Wales and adopted a stratified sampling design. The survey design over the 6 waves from 2002 to 2015 entailed home interviews with parents (main respondent and partners), testing of children (cognitive, physical and health measures), teacher surveys (when children attended school), and interviews with children.

5.3.2 Sample selection

We used data from all sweeps that were available so far. Children and their parents have been surveyed at child ages of 9 months (Sweep 1), 3 years (Sweep 2), 5 years (Sweep 3), 7 years (Sweep 4), 11 years (Sweep 5), and 14 years (Sweep 6). MCS Sweep 1 started with a productive sample of 18818 children in 18552 families. Additional 1389 families who fell into the English sampling frame but could not be reached in time were approached in Sweep 2. Overall 19,243 families and 19,517 children have ever participated at the MCS. From this sample, we removed first 198 children (192 families) for whom no information on analytical core variables such as parental education, household income, or ethnicity was unavailable. Second, since in the context our analysis essential information on families' migration experience had been collected from the second sweep, we conditioned our analytical sample to those children and families who participated at MCS Sweep 2 (15,440 families, 15,653 children). Third, in order to achieve a balanced sample, we restricted our sample further to families and children who had participated continuously from Sweep 2–6 (9,379 families, 9,509 children).

5.3.3 Independent variables

Highest parental education
We reconstructed highest parental education by the highest number of years either parent spent in fulltime education. The first valid value across all sweeps was used. Less than 12 years were classified as 'low' parental education, 13–15 years as 'medium', and 16 or more years as 'high'.

Economic position of household
We draw upon a derived variable that measured the income quartiles a particular family is looked at in the income distribution across the UK. The highest income value across the first three waves was used. In addition, we used a derived variable demarking family living below the poverty line (60% of the median income).

Migration background, ethnicity, and cultural proximity
We created a variable indicating whether children have a migration background. Having a migration background is defined as having at least one parent who was not born in the UK. Analyses further divide between three groups children with no immigrant parents, with one immigrant parent, and two immigrant parents. Note that by sample design first-generation migrant children are not included in the MCS. For child ethnicity, we drew upon a six-fold classification that had been provided by the MCS data and distinguished between white, mixed, Indian, Pakistani and Bangladeshi, Black or Black British, and other (including Chinese). Furthermore, we measured cultural proximity through language spoken at home (three categories English only, English and other, other language only; first value across first three waves) and religiosity (of the main respondent, overwhelmingly the mothers) entailing the four categories Christian, Muslim, other, and none.

Family demographics

In some analyses we utilized data on family demographics, including the number of children living in the household (one, two, three or more children, highest value across first three waves), mother's age at birth of the cohort child, and family type (two parents, single parent, other; last value across first three waves).

Table 2 provides an overview to characteristics of children and their families based for UK and separately for UK countries.

5.3.4 Child cognitive and educational achievement

To measure children's cognitive and educational achievement we relied on two sets of measures available in the MCS. First, we utilized score data from cognitive assessments taken at children's home. Table 1 provides an overview of measures taken at various ages. If available, we used ability or raw scores that were not pre-adjusted for age and other background characteristics (the MCS provides T-scores which are age adjusted scores). We then adjusted the raw test score data for age-differences by regressing test scores on child age using a quartic polynomial functional form and standardising the residuals from that regression to have a mean of 0 and standard deviation of 1. These *z-scores* measure in a standardised way variability in abilities of children that does not relate to age. Residualizing and standardization of test scores was conducted using attrition weighted data (see below).

In line with Hoffmann (2018) we classified the various assessments into two skill sets, 'verbal' and 'quantitative' skills. Verbal skills entail BAS Naming Vocabulary in Sweep 2 and 3, BAS Word Reading in Sweep 4, BAS Verbal Similarities in Sweep 5, and BAS Word Activity in Sweep 6. The second skill set, quantitative skills, encompasses the numbers/counting sub-scale of the Bracken School Readiness test at Sweep 2, BAS Pattern construction at Sweep 3, NFER Number skills and Math at Sweep 4, and the CANTAB Spatial Work Memory score (we created a reversed scale of the total number of errors) at Sweep 5. Unfortunately, no quantitative skill was available at Sweep 6.

Furthermore, we constructed a composite index for the multiple measures in Sweep 2 to 5. The score was calculated by averaging the z-scores for the test data in the respective wave. The average z-score was subsequently z-standardized again. We use the composite score as a summary measure for a child's position in the overall distribution of multiple cognitive domains.
The longitudinal analysis of gaps in cognitive achievement will be carried out using the composite index as well as verbal and quantitative skills separately. Employing the composite index has two advantages in our view. First, it takes into account all cognitive skills that have been assessed by the MCS sweeps. Second, by averaging results from several test, the composite index is in a way reducing potential issues related to measurement error in single tests (Feinstein 2003). Third, using a composite index is the only viable strategy to offer a broad overview on the evolution inequalities over the entire observation window (see chapter on Germany). However, as composite index relates to children's position in the overall distribution of multiple tests, it lacks specificity and, hence, may mask important gap heterogeneity across intellectual domains. This shortcoming is compensated by inspecting verbal and quantitative skills in addition to the composite index.

Table 1 Data on cognitive achievement in the MCS.

Child Age	Test	Description
3 years (Sweep 2)	*Bracken School Readiness*	Composite score of the Bracken School Readiness Assessment (BSRA). The Bracken Basic Concept Scale – Revised (BBCS–R) was used to assess the basic concept development in children. Concepts relating to colours, letters, numbers/counting, sizes, comparisons, and shapes were assessed in the BSRA.
	BAS Naming Vocabulary	Ability score that measures expressive language ability according to the British Ability Scales (BAS). Assessed was spoken vocabulary. Test items consist of a booklet of coloured pictures of objects which the child is shown one at a time and asked to name.
5 years (Sweep 3)	*BAS Naming Vocabulary*	See above.
	BAS Picture Similarity	Ability score that measures problem solving abilities. Children were shown a row of 4 pictures on a page and asked to place a card with a fifth picture under the picture most similar to it.
	BAS Pattern Construction	Ability score that measures spatial awareness, dexterity and coordination, but also perseverance and determination. Children had to construct a design by putting together flat squares or solid cubes with black and yellow patterns on each side. The child's score is based on accuracy and speed.
7 years (Sweep 4)	*BAS Word Reading*	Word Reading is an assessment from the BAS which assesses children's English reading ability. Child had to read aloud a series of words presented on cards (90 words in total with ascending order of difficulty). A small subset of Welsh children completed a Welsh language test instead of the English test.
	BAS Pattern Construction	See above.
	NFER Number skills and math	Score measuring children's mathematical skills (we used the raw not age adjusted score based on the original PM7 test). The test was adapted from the NFER Progress in Maths test which is aimed for 7-year-olds and was originally developed and nationally UK standardized in 2004.
11 years (Sweep 5)	*BAS Verbal Similarities*	Ability score measuring children's verbal reasoning and verbal knowledge. The interviewer read out three words to the child who had to say how the three things are similar or go together.
	CANTAB Spatial Work Memory	Score constructed by the number of errors across all SWM trials. Test assesses child's ability to retain spatial information and to manipulate remembered items in working memory.
14 years (Sweep 6)	*BAS Word Activity*	Activity score out of 20 words. The assessment involved presenting the respondent with a list of target words, each of which had five other words next to them. The respondent had to select, from the five options, the word which meant the same, or nearly the same, as the target word. The assessment measured respondent's understanding of the meaning of words.

Sources: Hansen (2014) and Fitzsimons (2017).

5.3.5 Teacher evaluations of children's performance in school subjects

In addition, to the cognitive assessments we exploited data obtained from the Teacher Surveys at age 7 and 11. Schools of children were identified, and teachers were asked to fill out a questionnaire assessing among other things children's progress in subjects such as reading and writing in English, Math, Science, or music and arts. For each subject, teachers evaluated a child's performance on a five-point scale that ranged from 'well below the average' to 'well above the average'. The data was only available for a subset of children there the analyses of teacher ratings was restricted to a sub-set of our analytical sample.

In our analysis we used teacher evaluation on 5 items in Sweep 4 (speaking, reading, writing in English, science, and math) and 3 items in Sweep 5 (English, science, math). For both Sweeps principal component analysis (PCA) was used to factor score the items. The PCA yielded a very clear one component solution for each sweep (Sweep 4: Eigenvalue 4 and 80 per cent item variance explained; Sweep 5: Eigenvalue 2.6 and 87.3 per cent item variance explained). Component scores were z-standardized to have a mean of 0 and a standard deviation of 1. Due to non-availability of teacher data the sample size reduced to 5,989 children for teacher evaluations in Sweep 4 and 5,426 children for evaluations in Sweep 5. Inversed probability weighting was used to calculate special weights that adjust the attrition weights to account additionally for selective attrition in the teacher evaluation data.

5.3.6 Missing data

79% of the 9509 children of our sample had complete data on all cognitive assessments listed above. We used multiple imputation to account for missingness in 21% of the cases. We used chained equations allowing for mixed distributions and drew 5 imputations from the approximated joint distribution of variables. The imputation model included all our analytical variables and ancillary variables.

5.3.7 Accounting for attrition

Selective attrition and participation in the MCS survey may induce a selection bias in our balanced sample. Especially harmful to our longitudinal analysis were participation and attrition being a function of cognitive and educational achievement of children. For instance, if low performance on test was increasing the probability of non-participation our estimates of the development of achievement gaps across social and ethnic groups could seriously suffer from a positive selection bias leading to a flawed estimation of actual gaps most likely an underestimation. Although the MCS group provide longitudinal weights accounting for attrition, these weights implement more a general-purpose strategy and, unfortunately, do not account for the possibility of attrition being selective by child achievement. Therefore, we decided to construct our own longitudinal weights through a procedure that is informed by our analytical aims.

For constructing attrition weights, we started with the full sample of children that participated at Sweep 2 (not the balanced sample). For all these children we used logit models predicted the probability of participating at Sweep 3. To address non-monotone dropout patterns in the data, i.e. families who participated at Sweep 2 skip participation for example at Sweep 3 but return to the panel in Sweep 4, we additionally predicted probabilities to participate at Sweep 4, 5, and 6. Analogously, conditional on participation at Sweep 3, we predicted the probability of participation at Sweep 4, 5, and 6, and conditional on Sweep 4, participation at 5 and 6 an so forth. The logit models included all analytical variables – independent variables and cognitive assessment variables – and also ancillary variables. Missing data in the cognitive assessments were imputed sweep-wise via multiple imputation (five imputations). Logit models were estimated for all imputation datasets and average predicted probabilities were used for weighting. Weights were then constructed by multiplying the initial weight in Sweep 2 (provided by MCS) with the inverse of the path-dependent participation probabilities. For example, the weight in Sweep 6 for a child who participated throughout all Sweeps is w_2 x $1/p_{32}$ x $1/p_{43}$ x $1/p_{54}$ x $1/p_{64}$, and for a child who skipped Sweep 3 and 4 but reappeared in Sweep 5 and 6 the weight was w_2 x $1/p_{52}$ x $1/p_{65}$. Finally, we matched all Sweep 6 weights to the balanced sample and used these weights for all analyses.

Estimates from the logit models predicting participation indeed showed evidence for selective attrition by children's cognitive achievement (see Appendix 5.1). Hence, we believe that for our purposes our weighting approach is superior to the standard weights as provided by the MCS data. Table 2 provides sample estimates on the distribution of selected variables among children in the UK as well as separated by country.

Table 2 Children born 2000–2001 in the UK.

	UK	England	Scotland	Wales	North. Ireland
Female (%)	49.5	49.6	49.0	47.9	49.9
Parental education (%)					
low	51.7	51.2	56.8	52.8	48.0
medium	25.2	25.6	19.7	25.7	27.4
high	23.2	23.2	23.6	21.5	24.6
HH net income (weekly) (mean)	604	609	606	555	556
Poverty (below 60% median) (%)	17.5	17.7	14.3	19.6	19.4
Migration background (%)	14.8	16.8	5.2	7.0	4.2
White ethnicity (%)	86.4	84.0	98.0	97.1	99.5
Children in HH (mean)	2.4	2.4	2.3	2.4	2.6
Two parents in HH (%)	76.2	76.1	78.0	74.0	76.5
% of all children	100.0	82.8	8.6	4.9	3.6
Case numbers	9509	6221	1031	1350	907

Notes: Balanced sample. All data weighted. Case numbers unweighted.

5.4 Methods

5.4.1 Studying the evolution of achievement gaps from preschool to school age

To address the first set of questions, we will first study the evolution of gaps by estimating distances between the average z-scores among groups for the three outcomes (composite, verbal skills, and quantitative skills). Furthermore, we will apply stepwise multiple linear regression models to assess how estimated group differences change once we adjust for the effects of other covariates. We applied multivariate models to better understand the structural mechanisms of achievement gaps by parental education as our prime SES indicator and migration background. For instance, for studying migration gaps, we start with a baseline model estimating overall group-inequality in z-scores

$$z_{ti}^{k} = a_t + b_1 MIG_{1i} + b_2 MIG_{2i} + \varepsilon_i \quad (1)$$

with index *i* denoting a child, index *t* denoting the age stage (Sweep), and superscript *k* denoting the outcome (composite, verbal skills, and quantitative skills). MIG denotes dummy variables for the categorical variable migration background (0 = no migration background, 1 = one immigrant parent, 2 = two immigrant parents). Coefficient $b_1$estimates then the z-score gap between children with one immigrant parent and children without migration background. Correspondingly, coefficient b_2 estimates the gap in z-scores that separates children with two immigrant parents and children without migration background. We will refer to that as the 'overall' gaps.

In a second step, we included, in a stepwise fashion, additional covariates to adjust the gaps estimation for factors that are associated with migration background but that are also predicting achievement. The models that the following form:

$$z_{ti}^{k} = a_t + b_1 MIG_{1i} + b_2 MIG_{2i} + \mathbf{Cz} + \varepsilon_i \quad (2)$$

with **C** being a vector including additional covariates and **z** being a vector with associated coefficients. For interpretation, it is essential to note the difference in the theoretical meaning of the migration coefficients b in Equation 1 and Equation 2. While Eq. 1 coefficients measure the *overall gaps* related to migration status in children, Eq. 2 coefficients measure the *adjusted gaps* that is the residual gaps related to migration status that remain after we factor out the influences of covariates **C**. In other words, these are gaps between migration gaps that remain unexplained after accounting for covariates **C**. For instance, on average, migrant parent households earn less than households with native parents. Based on theories of family investment (Conger, Conger, & Martin 2010; Duncan, Magnuson, & Votruba-Drzal 2015), income represents an important indicator for the ability of parents to invest in their children. In fact, empirical research has shown that families income levels predict child achievement (Sirin 2005). Consequently, migrant gaps in child achievement at least in part might be attributable to the fact that migrant families possess less financial capital being available for child investments. A comparison of overall and adjusted gaps between groups provides further insights as to which extent covariates explain those overall group gaps.

5.4.2 Estimating the additional influence of background characteristics during school

To address the second set of research questions for the UK we adopt for the same strategy as in the chapter on Germany. That involves analysing the effects of SES and migration background on cognitive achievement in school ages by conditioning on earlier achievement levels when children entered school. Based on the logic of a standard mediation analysis we aim to estimate how much of the group-based inequality in school-age abilities is attributable to group-based inequalities when children just entered school age. The indirect association of group membership and school-age achievement that runs through the association of earlier with later achievement levels (henceforth ‘indirect effect’). The direct association of group membership and school-age achievement is the association that remains after accounting for earlier achievement (henceforth ‘direct effect’). Both indirect and direct effect sum up to the total association between group membership and school-age achievement (the gaps studied in the first step), which we call henceforth the ‘total effect’.

At the example of migration related gaps, the structural model estimating the direct effect of migration background is defined as

$$z^k_{t>5,i} = a_t + b_1 MIG_{1i} + b_2 MIG_{2i} + b_3 z^k_{t=5,i} + \varepsilon_i \qquad (3)$$

with $z^k_{t=5,i}$ denoting a child’s achievement level at age 5 and $z^k_{t>5,i}$ denoting a child’s achievement level at later ages in school life (7, 11, and 14). Coefficient b_3 is an estimate for the average associational persistency in achievement and coefficients b_1 and b_2 represent the direct effects of the two categories of migration background versus the reference of no migration background. This contrasts to the total effects measured in Equation (1). In a second step, we condition the mediation model on the full set of control variables

$$z^k_{t>5,i} = a_t + b_1 MIG_{1i} + b_2 MIG_{2i} + b_3 z^k_{t=5,i} + \mathbf{Cz} + \varepsilon_i \qquad (4)$$

Note using Equation (3) to estimate the link between earlier and later achievement involves issues of endogeneity if error terms of $z^k_{t=5,i}$ and $z^k_{t>5,i}$ are correlated. A correlation of errors can arise first of all from an omitted variable bias. Although, we include a series of control variables in (4) it is likely that earlier and later achievement levels are jointly influenced by unobserved heterogeneity (e.g., the genetic endowment of children). Hence, we interpret *b3* in (3) and (4) as a partial association predicting later achievement based on earlier but not as an estimate for the causal effect of earlier on later achievement.

Second, endogeneity issues arise also in relation to the accuracy by which true achievement levels $z^k_{t=5}$ have been captured by the MCS tests. Measurement error in the test scores on the right-hand side of the equations create endogeneity leading to regression-to-the-mean effects. For instance, a child may have been lucky (unlucky) to score extraordinarily high (low) in the test at age 5. Being lucky or unlucky is not part of the true ability but enters the test score as measurement error. Consequently, on the next test occasion the previously lucky (unlucky) child is expected to score lower (high) at a value closer to the true ability as the child tends to regress to its mean performance. Regression to the mean imposes challenges to empirical estimation of (3) and (4) based on observed scores. If measurement error in test scores at age 5 is non-negligible coefficient b_3 will be downwardly biased and b_1 and b_2 upwardly biased. While the estimation of the total effect of migration background is not affected by measurement error, we tend to overestimate the direct effect and underestimate the indirect effect.

To minimize bias from measurement error in estimation of Equations 3 and 4, we applied instrument variable (IV) estimation which is a commonly established solution in longitudinal research on achievement (Bradbury et al. 2015). The basic idea being using IV estimation is to take another achievement score (the instrument) for predicting age 5 scores. These predictions represent an error free signal of age 5 ability and are plugged into (3) and (4) instead of the measured (error prone) scores. Our IV strategy exploits earlier measurements at age 3 to instrument age 5 ability scores. Instruments have been selected depending on the type of outcomes. Values of the composite ability index at age 5 were instrumented by composite ability at age 3, quantitative scores at age 5 by scores in the number/counting sub test of the Bracken school readiness test at age 3, and verbal skills at age 5 were instrumented by the BAS naming vocabulary test at age 3. We adopted two-stage least squares estimation using *ivregress 2sls* in Stata. The first stage equation predicting age 5 scores included age 3 scores in quadratic functional form and all other covariates from the second equation (aiming to estimate models 3 and 4).

Finally, we study how SES or migration effects differ along the distribution of earlier achievement (also see chapter on Germany). To address these questions, we adopted a similar approach to the one proposed earlier by the comparative study from Bradbury and colleagues (2015) who called this the 'divergent trajectory' model.[1] For that we estimate additional models based on Equation 3 that include an interaction between age 5 performance and group membership. For example, the interaction model for migration background is

$$z_{t>5,i}^{k} = a_t + b_1 MIG_{1i} + b_2 MIG_{2i} + b_3 z_{t=5,i}^{k} + b_4 \cdot MIG_{1i} \cdot z_{t=5,i}^{k} + b_5 \cdot MIG_{5i} \cdot z_{t=5,i}^{k} + \varepsilon_i \qquad (5)$$

Using Equation 5 we will simulate group-specific trajectories starting from different levels of initial age 5 performance.

5.4.3 Accounting for complex survey design

All analyses accounted for the complex survey design of the MCS data which have had been collected by stratified and multi-stage cluster sampling. Moreover, all analyses are weighted by using the weights calculated for the balanced sample. All estimates on cognitive achievement were run on the five imputation datasets and estimates and standard errors were combined accordingly (Heeringa, West, & Berglund 2010).

5.5 Results

5.5.1 Achievement gaps: An overview

Figure 1 provides a compact overview to gaps by a series of individual and family characteristics in cognitive achievement as measured by the composite ability index (see Appendix 5.2 for detailed estimations). We first note sizeable achievement gaps along several dimensions. On average we find that girls perform better than boys, children from families with higher SES (in terms of parental education and income) perform better than children from lower SES families, children without migration background perform better than children with migration background, children of families with home language English perform better than children from homes that speak other languages, and white children perform better than children from other ethnicities. Gaps are largest at the first observation of age three and for the most part shrink in the

[1] Note that for their UK analyses, Bradbury and colleagues (2015) used the same MCS data inspecting divergent trajectories by parental education. Our estimates presented here in part replicate their findings (for age 5 to age 11, the range used by Bradbury and colleagues) very well and estimates are very similar. Small deviations are explained by a different sample cut, a different classification for educational levels, different weighting strategy, and subtle differences in the IV estimation. Our analyses on divergent trajectories by parental education extend earlier estimates by including age 14 and providing additional estimations for quantitative skills and the composite ability index.

subsequent years. For instance, at age three the poverty gap in cognitive achievement is substantial with about 85% of a SD lower achievement for children from income poor families, a gap which equals roughly four times the gender gap in achievement. As children grow older, however, the poverty-achievement gap is shrinking and is almost cut by half by age 14 albeit still remaining statistically significant and with a difference of about 45% of a SD substantial. Concerning parental education, however, gaps remain by and large stable over age stages with an average difference of 60 to 70% of SD for children from high versus children from low educated parents.

While children of immigrants and children of minority ethnic groups are disadvantaged in the early years, those inequalities are largely vanishing as children progress through school life. While children of immigrants lag behind by roughly 60% of a SD in age 3 abilities, they are gaining substantial grounds as time passes by. At age of 14, migrant children and children of natives are indistinguishable in cognitive-academic performance. Almost the same holds true for the gaps between white majority children and children from ethnic minority groups; gaps are sizable in the early years (80% of a SD, age 3), but substantially decline over course of schooling although a small 'white' advantage remains. We find analogous results for language culture at home.

In Figure 2 we see the same overview for compound scores of teachers' evaluation of children's achievement. Although teacher evaluations are restricted to age 7 and 11, inspecting gaps in teacher evaluations yield conclusions pretty similar to cognitive achievement scores. Some few inconsistencies are visible with respect to gender and ethnicity. While girls' advantage in cognitive achievement disappeared in cognitive skills by age 11, girls are still better evaluated by teachers. As regards ethnicity, we note no (aggregate) minority disadvantage in teachers' evaluations at age 11 whereas we did find a minority disadvantage in cognitive-achievement at the same age. Possible explanations for these discrepancies between cognitive-achievement and teachers' evaluation of educational achievement may (among others) include be an imperfect association between cognitive achievement as assessed by the MCS home tests and students' performance in school, the role of other skill sets like non-cognitive skills related to class room behaviour which might enter teachers' evaluations, or a teacher bias in evaluation. Overall, however, both perspectives – cognitive achievement and teacher evaluations – reveal more homogeneity rather than heterogeneity in findings. Furthermore, scores of the composite ability index were highly predictive for teacher evaluations (correlation of .70 at age 7 and .58 at age 11). Based on that, we restrict all following analyses to cognitive-achievement scores.

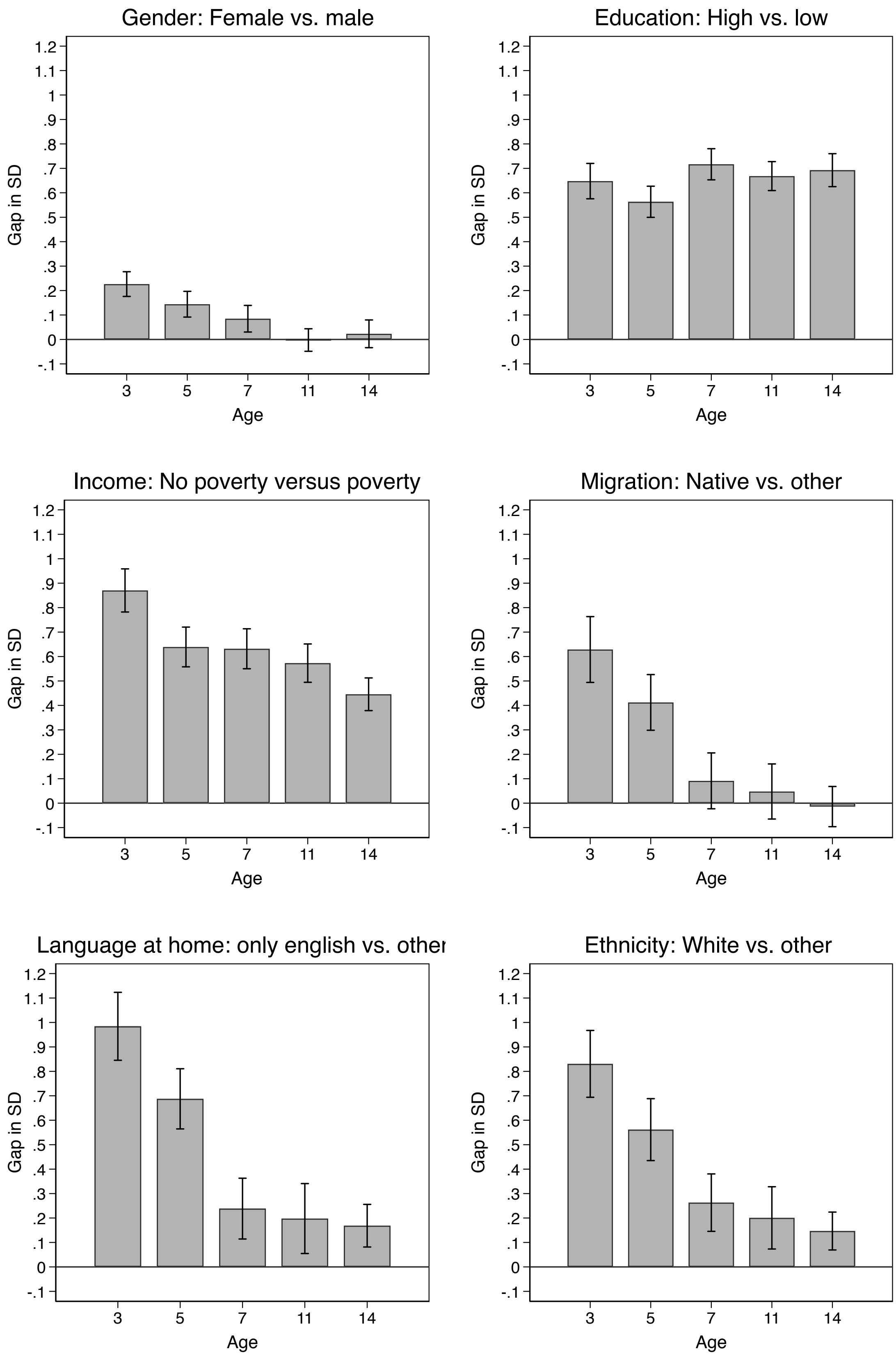


Figure 1 Gaps in cognitive achievement across age by various background characteristics.

Notes: Composite index (age 3 to 11), Word reading (age 14).

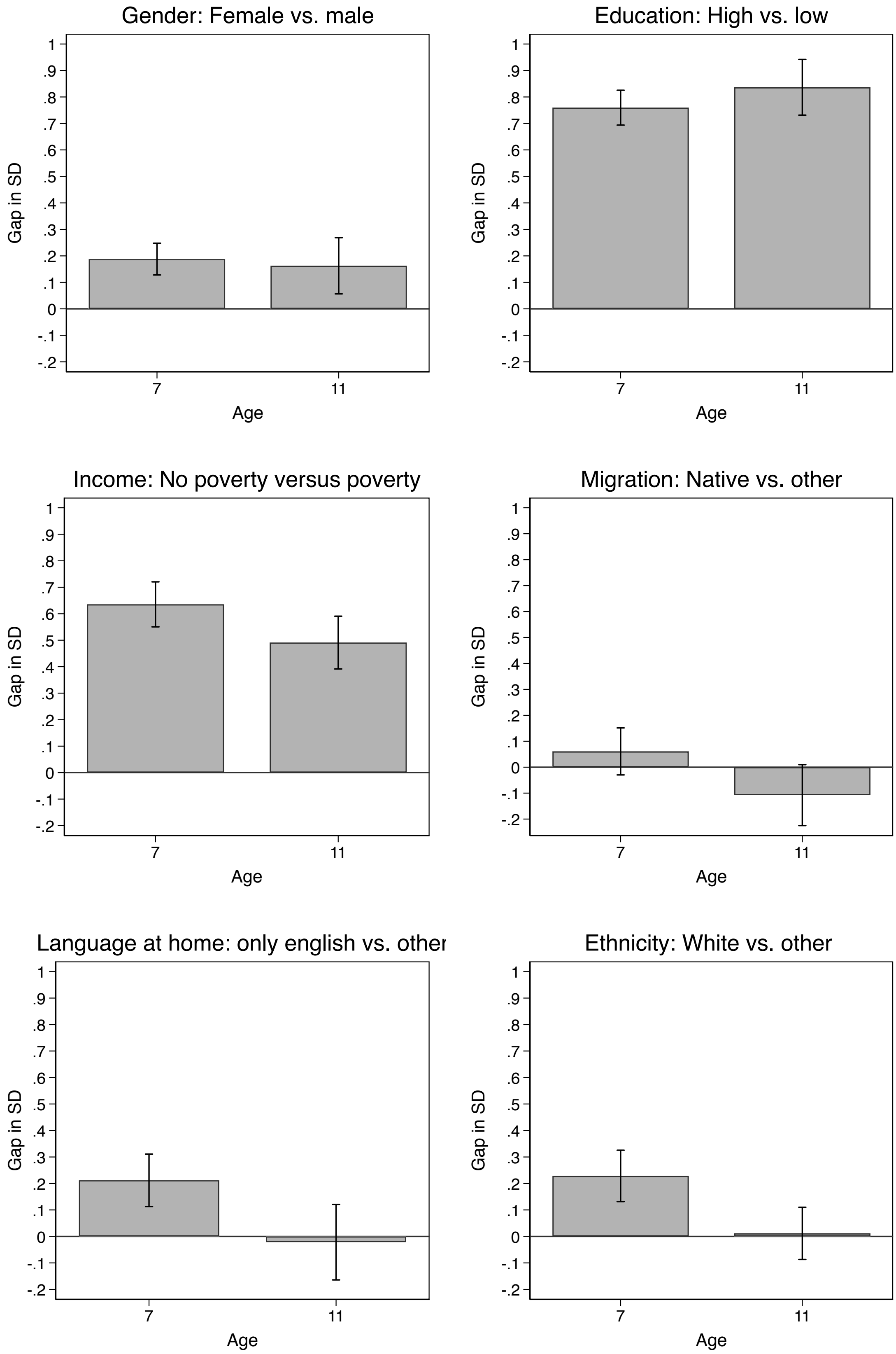


Figure 2 Gaps in teacher evaluation of academic achievement across age by various background characteristics.

Notes: Teacher evaluations in Sweep 4 and 5 (standardised component scores).

5.5.2 Evolution of gaps in cognitive achievement from early years to secondary schooling

In the following, we will provide a detailed report on the evolution of gaps in cognitive achievement from age 3 to 14 along our two principal analytical axes socio-economic status on the one hand, and migration, ethnicity and minority status on the other hand. Analysed are gaps in three achievement outcomes; the (a) composite ability index, (b) verbal skills, and (c) quantitative skills. We start by analysing overall gaps between groups (unadjusted mean differences across groups). In a second step, we add multivariate perspectives that provide deeper insights as to which degree socio-economic and ethnic achievement gaps remain when adjusting for covariates.

5.5.2.1 Socio-economic status

How large are social gaps in achievement and how do they develop from preschool to school age?

Figure 3 provides evidence for the question at hand. Plotted are outcome levels by parental education, income quintiles, and the binary poverty status grouping. For having a compact illustration, we graphically overlaid estimations for composite index (bold black), verbal (rose) and quantitative skills (light blue). Concerning the composite ability index, we see – by and large – stable gaps by parental education. Although there are some minor fluctuations, the high-low gaps at age 11 and 14 are statistically not different to the gap at age 3. However, income-achievement gaps are shrinking at the same time: the gap between children from the top and the bottom income quintiles was 1.24 SD at age 3 but subsequently shrunk to .87 SD at age 11 and .76 SD at age 14 (reductions were statistically significant). Hence, although income-achievement gaps were large before children entered school, gaps were continuously decreasing after children transited to school life. Note that the decrease in inequality was most pronounced between the bottom and middle-income classes, while the top income class was gaining ground over middle classes. The third panel (at the bottom) finally provides a more focused comparisons of children from impoverished versus not impoverished families. Similar to our observation with regard to income classes, the poverty gap shrunk from .87 SD at age 3 to .57 SD at age 11 and .44 SD at age 14.

We get a more differentiated picture when singling out verbal and quantitative skills. Parental education gaps in verbal skills first increase from age 3 to 7 but remain stable afterwards. In contrast, income and poverty gaps in verbal skills were shrinking consistently. Interestingly, socio-economic gaps in quantitative skills – although when compared to verbal skill gaps being smaller in size – were by and large stable across the whole observation window. Taken together, achievement differences between SES groups are substantial throughout children's preschool and schooling careers even though income gaps tend to reduce. No conclusion changes when we adjust these figures for migration background (see Appendix 5.3).

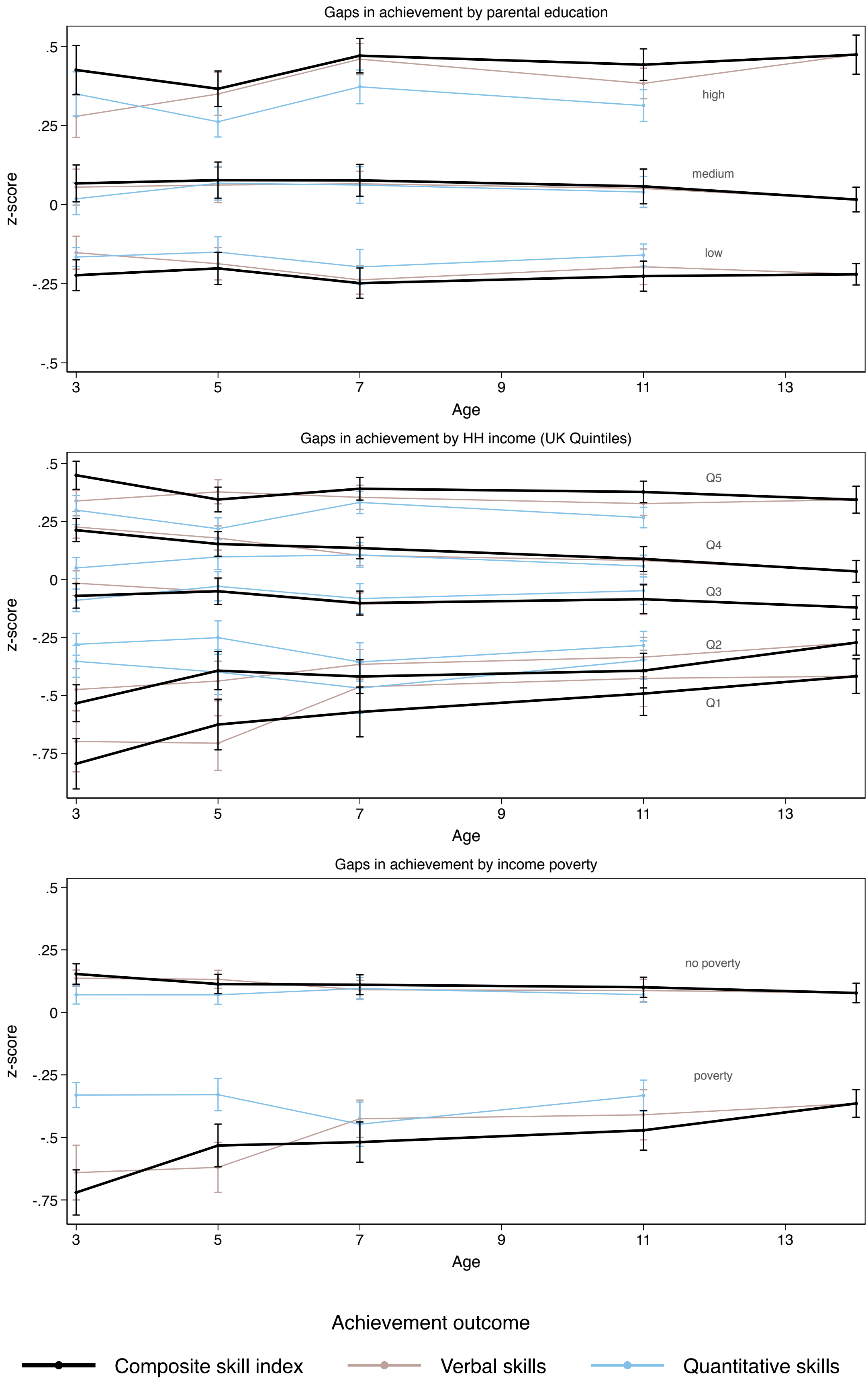


Figure 3 Evolution of SES-achievement gaps.

5.5.2.2 Migration background and minority status

How large are achievement gaps by migration background and minority status and how do these gaps develop from age 3 to 14?

To address this question, we will present estimates on achievement gaps by (1) migration background, (2) child ethnicity, and (3) religious orientation in the home. We took several indicators into account because all of them mark minority status but highlight potentially different facets which are likely to correlate with differential experiences at home and in school. Table 3 shows some descriptive figures about the prevalence of children with minority background in the UK context estimated based on our sample. About 15% of children had a migration background measured by having at least one immigrant parent (see Table 2). A portion of 26% of migrant children had two immigrant parents. Moreover, about 14% of the children were of other than white ethnicity with Pakistani and Bangladeshi ethnicity the model category (32% of non-whites) among them (followed by mixed ethnicity and black). In terms of religious affiliation, a majority of children's mothers were Christian (46%) or had no affiliation (46%). A minority of children had Muslim mothers (6%) or mothers of other beliefs (2%) such as Hinduism, Sikhism, Judaism, and Buddhism.

In Figure 4 we present the predicted achievement levels by the different minority dimensions of migration background, ethnicity, and religious culture, separately for each outcome. For sake of readability, confidence intervals were left out from the figure (but are provided in the Appendix 5.4). In the early age stages, we find that children of immigrants (upper panel) substantially lag behind the children of natives, with the largest disadvantage being visible for children of two immigrant parents. For instance, the gap in composite ability between children with two-immigrant parents and children with native parents amounts to 1.08 SD at age 3 (and, .48 SD for one immigrant parent versus native parents). However, the immigrant kids' relative disadvantage is thawing rapidly over subsequent years. By age 7, after 2 years of schooling, the two-immigrant-vs-natives gap reduced to .16 SD ($p < .05$) and the one-immigrant-vs-natives gap close to zero (.07 SD, $p = .268$). By age 11 and 14, none of the group differences are statistically significant anymore. As the distinction by skill type shows, the shrinking gaps are mainly explained by an astonishing convergence in verbal skills. In the English reading test at age 7, immigrant children even outperform children of native British by 13 to 17% of a SD ($p < .05$ for both comparisons). Group differences in verbal skills at age 11 and 14 are small and lack statistical significance.

Findings differ starkly when looking at quantitative skills. Although differences are way smaller compared to composite and verbal skills (e.g., two-immigrant vs native gap = .16 SD, $p < .05$), they are astoundingly persistent and not reducing over early school years (although differences are not significant by age 11).

Ethnicity (central panel) does matter for early cognitive achievement but – alike for migration background – the ethnicity-achievement association is getting weaker over schooling. Children being Pakistani or Bangladeshi stand out by facing the largest relative disadvantage when looking at composite and verbal skill scores. At age 3, a massive verbal skill gap of 1.62 SD ($p < .001$) separates Pakistani or Bangladeshi from white majority children. Although reducing this gap remains substantial and significant at ages of 11 (.52 SD) and 14 (.33 SD). But ethnic minority is not always related to a gross disadvantage. For example, Indian children are occasionally outperforming white children in school (particularly at age 7 and 11). Ethnic disparities are comparably small and more stable for quantitative skills for which we found no clear evidence for differences between white, mixed ethnicity, Indian, or ‘other’ ethnicity children. The groups of Black and Pakistani/ Bangladeshi children, however, face significant disadvantages compared to the other groups. Again, largest disadvantage can be seen for Pakistani/Bangladeshi children. Depending on age they are lagging .33–.65 SD behind the white majority group.

Before interpreting the gaps by religious affiliation (bottom panel in Figure 4), one should keep in mind that there is a substantial intersection of migration status, ethnicity and religious belief. As one can learn from Table 3, the two – in terms of size – dominant groups of Christian and None are overwhelmingly made up by white children with only very little migration background. In contrast, three quarters of children with Muslim mothers are ethnic Pakistani or Bangladeshi and nearly 90% have a migration background. The smallest group ‘other’ religion is more mixed with about 60% Indian ethnicity and nearly 70% migration background.

Bearing those compositional differences in mind, we can conclude that little gaps are visible between Christian, ‘Atheists’, and others. Muslim children, however, face consistent disadvantage in all outcomes whereas disadvantage is most pronounced in verbal skills and least pronounced in quantitative skills. That disadvantage in verbal skills is large in the early years and then diminishes over school age, although the difference to the average remains statistically and substantially significant (e.g., gap between children from Christian and Muslim mothers is 1.56 SD at age 3, 1.42 SD at age 5, .12 SD at age 7, .53 SD at age 11, and .33 SD at age 14). Instead, the lower differences in quantitative skills between Christians and Muslims are more constant over the early life course, although a similar pattern of decreasing inequality can be found over the school years (.62 SD at age 7, and .37 SD at age 11).

Furthermore, Figure 5 reveals that migration background and ethnic minority status interact (find specific estimates and confidence intervals in Appendix 5.5). As we can see, while white-non-white differences are small for children without migration background, they are larger for children with migration background.

In conclusion, we can say that studying minority-majority achievement gaps invokes multiple perspectives as there are subtle difference in findings depending on the viewpoint. In general, we can conclude that minority-majority achievement gaps are initially large especially in language skills but – by and large – disappear when children navigate through school life. Hence, overall, we may conclude that although many minority children start with a relative disadvantage, they are able to catch up before long. In contrast, gaps in more abstract, quantitative abilities are small but also more robust. Finally, there are certain ethnic and cultural groups that face a particularly strong disadvantage. Without doubt, social policy should be concerned about the situation of these children. On the other hand, the evidence presented here has shown that there are also minority groups that perform equally or even better than the majority group.

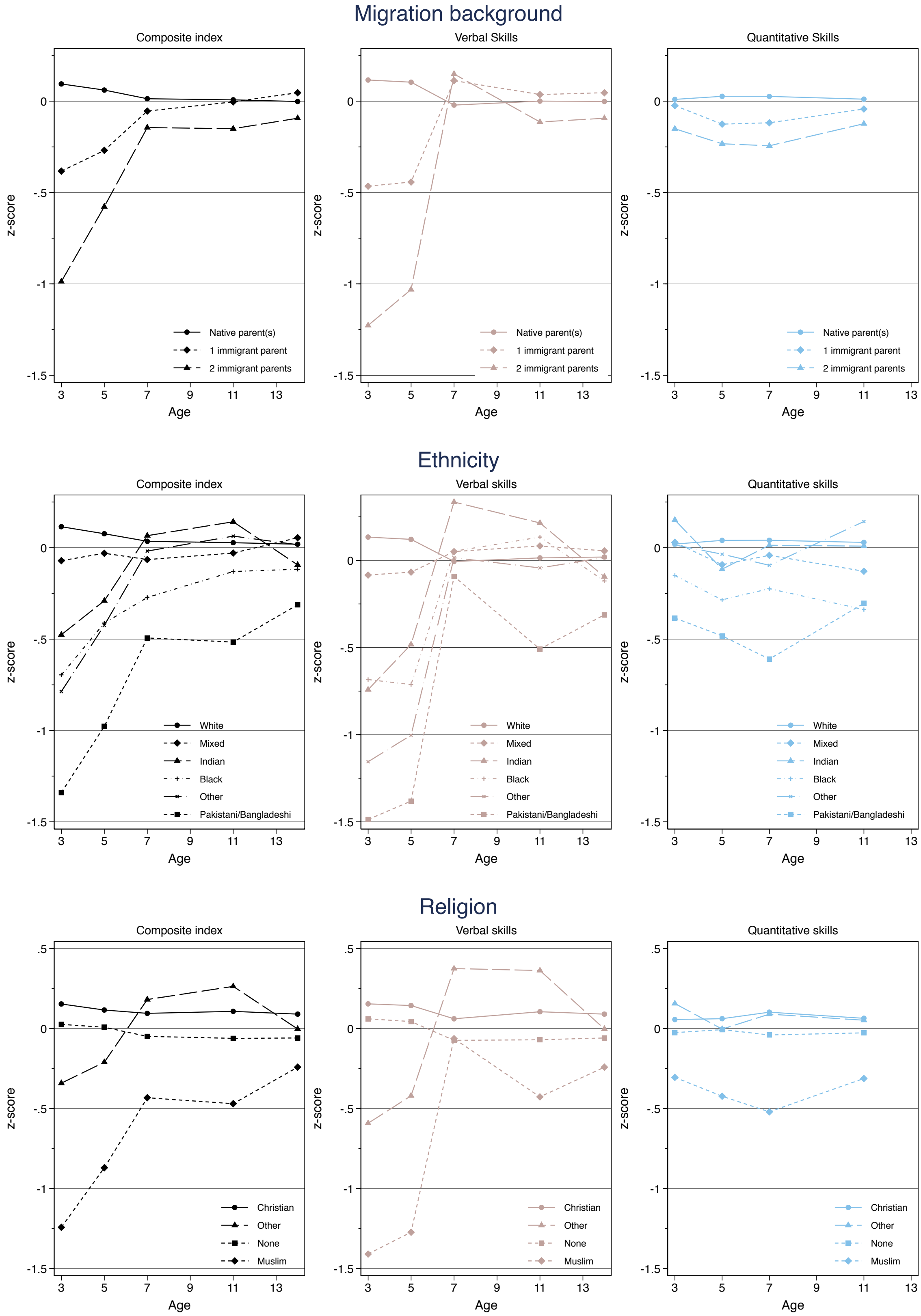


Figure 4 Achievement levels by minority status (migration background, ethnicity, religion).

Table 3 Ethnicity, migration background, and religion (column %).

	Religious group				
	Christian	Muslim	Other	None	Total
Ethnicity					
White	94.3	1.9	15.6	95.1	84.9
Mixed	2.1	3.7	3.4	3.0	2.6
Indian	0.2	5.1	59.2	0.3	2.6
Pakistani or Bangladeshi	0.0(4)	76.1	4.3	0.3	6.1
Black or Black British	3.1	7.3	1.2	0.8	2.4
Another ethnicity (including Chinese)	0.3	5.8	16.3	0.5	1.4
Total	100.0	100.0	100.0	100.0	100.0
Migration background					
No migrant parent(s)	89.8	11.8	31.9	93.2	83.3
One immigrant parent	8.7	50.9	28.2	5.9	11.4
Two immigrant parents	1.4	37.3	39.9	0.9	5.3
Total	100.0	100.0	100.0	100.0	100.0
Relative group size (row %)	48.7	7.6	3.4	40.3	100.0

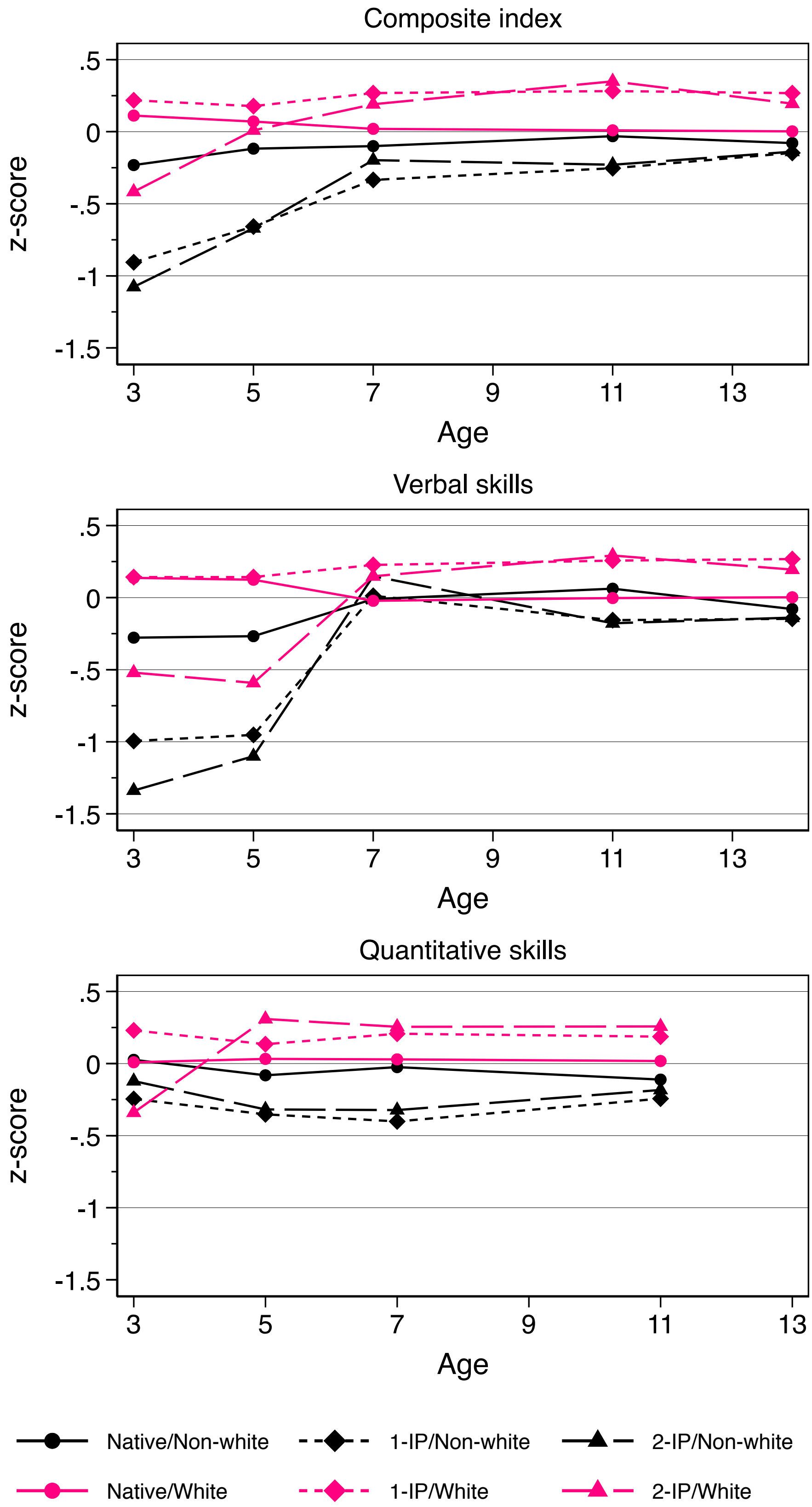


Figure 5 Achievement gaps by migration and ethnic minority status (non-white vs. white).

Notes: 1-IP = One immigrant parent. 2-IP = Two immigrant parents. Native = no immigrant parents.

5.5.2.3 Multivariate models of cognitive achievement

In this section, we present findings from multivariate models. Following the logic of Equation 2 (see the method section), we estimate achievement gaps by adjusting them for other covariates in a step-wise fashion. We present adjusted versions of achievement gaps for the two main stratifying dimensions parental education and migration background.

We begin with gaps by parental education. As Table 4 shows, parental education is associated with several other factors of advantage (e.g., higher income, more stable families, less teenage mothers) and disadvantage (children of higher educated parents more frequently have migration background or belong to an ethnic minority group at the same time), so the overall gaps by parental education represent the compound of a complex operation of various mechanisms and compositional differences across groups. Hence, to get a better estimate for the direct association of parental education and children's performance at various ages – after other important associated influences have been factored out – we present adjusted gaps by parental education. This allow us to understand the extent to which these characteristics explain achievement gaps by parental education.

First, we estimate overall gaps taking high parental education as reference category. The corresponding gaps – low versus high and medium versus high – equal the difference between the respective curves presented in Upper Panel of Figure 3. In a second model, we controlled for migration background. The third model additionally controls for child ethnicity and language spoken at home. The fourth model adds variables measuring family income (quintiles). Finally, in a fifth model, we added family characteristics (number of children in the household, disrupted family status, and mother's age at child' birth) and child's sex. All five models have run on the composite index (five age stages), verbal skills (five age stages), quantitative skills (four age stages). Hence in total, we estimated 65 models (available upon request to the authors).[2] The main findings obtained from this exercise are summarized in Figure 5. Full model specifications (Model 5) are shown in Table 6 (composite ability index), Table 7 (verbal skills), and Table 8 (quantitative skills).

Black lines represent the overall gaps that have been studied earlier. Once adjusting for compositional differences in migration experiences and culture (Models 2 and 3), gaps get slightly more pronounced in the earlier years. A reason for this is compositional differences between migrant and non-migrant families creating to counteracting effects: migrant families are characterised by higher parental education which acts as a positive resource for children but at the same time migrant status is related to lower achievement of children. Differences in family income, however, is one of the major factors explaining achievement gaps by parental education (Model 4). Residual gaps do not change after additionally accounting for individual and family characteristics (Model 5). After adjusting for all other covariates (Model 5), the relative disadvantage (in the composite ability index) for children from low educational background compared to children from high educated reduced by 46% at age 3, by about 39% for ages 5 to 11, and by 28% at age 14. Residual gaps by parental education remain substantial, statistically significant, and overall persistent over the early life course of children (e.g., about .5 SD for low

[2] It was 65 instead of 70 models, because the composite index at age 14 uses the same scores as verbal skills at age 14.

versus high contrast at age 14; for details see Tables below for full specification). The findings for adjusted gaps by parental education are very similar for all outcomes under study.

Table 4 Economic, cultural, and demographic composition of educational groups.

	Parental education		
	Low	Medium	High
HH Income quintile			
1st	13.8	4.8	1.7
2nd	27.4	13.2	5.5
3rd	23.8	21.1	8.2
4th	20.4	30.7	20.1
5th	14.5	30.3	64.6
Migration background			
No immigrant parents (native)	88.8	84.9	77.3
One immigrant parent	8.4	11.0	16.5
Two immigrant parents	2.8	4.1	6.2
Ethnicity			
White	88.4	85.4	83.0
Mixed	3.0	3.3	3.9
Indian	0.9	2.1	3.5
Pakistani or Bangladeshi	4.2	5.3	3.7
Black or Black British	2.5	2.7	3.9
Another ethnicity	1.0	1.2	2.0
Language spoken at home			
Only English	92.3	88.9	85.1
English in part	5.6	8.7	12.1
No English	2.1	2.5	2.8
Siblings at home			
None	17.7	16.5	12.4
One	41.8	51.5	54.1
Two or more	40.5	31.9	33.5
Family type			
Two parents	66.8	81.7	91.2
Single mum/dad	26.5	14.3	7.4
Other	6.7	4.0	1.4
Mother's age at birth			
20 or younger	17.1	7.2	0.8
21–25	20.0	20.1	7.6
26 or older	62.9	72.8	91.6
N	4342	2548	2619

Notes: All data weighted. Case numbers unweighted.

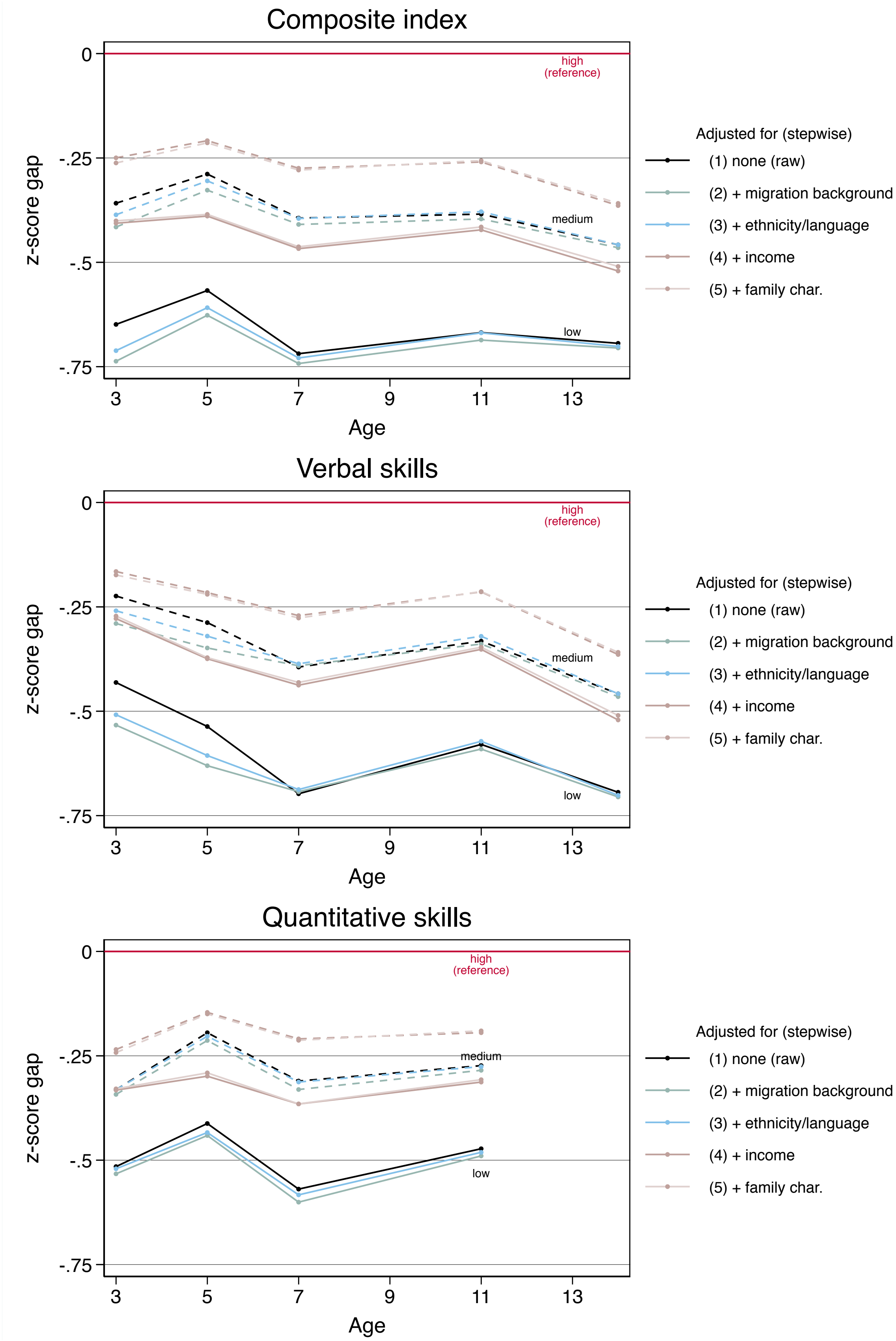


Figure 6 Achievement gaps by parental education step-wisely adjusted.

Notes: 'high' parental education is the reference (red line).

We carried out the same analyses for adjusting achievement gaps by migration background. Table 5 shows the compositional differences across groups of migration status. As we can see children with migration background tend to have higher educated parents, more frequently grow up in income poor families, more frequently belong to ethnic minority groups, have less English language exposure at home (particularly children with two immigrant parents), and have larger number of siblings but also live less often in disrupted families. Taken together, there are important compositional differences between children from immigrant and native families in various variables that might be related to advantage and disadvantage in terms of achievement. Hence, by a stepwise regression we attempt to better understand how much of the migrant advantage or disadvantage remains once we account for compositional differences between groups. We applied the same estimation strategy as for parental education, however, for theoretical reasons changed the order of variables being included: unadjusted gaps (Model 1); adjusted for socio-economic factors parental education and income (Model 2); additionally, adjusted for English language spoken at home (Model 3), ethnicity (Model 4), and family characteristics and sex (Model 5). Again 65 models had been estimated (available upon request to the authors) and findings are summarized in Figure 6. Full model specifications (Model 5) are the same as for parental education and are shown in Table 6 (composite ability index), Table 7 (verbal skills), and Table 8 (quantitative skills).

As visible from Figure 7, compositional differences by SES (Model 2) explain little (two immigrant parents versus native) to none (one immigrant parent versus native) of the gaps in the composite index of achievement between children with migration background and without. However, when looking at the composite index, gaps are substantially reduced when adjusting for language spoken at home (Model 3). Gaps slightly reduce further when adjusting for the different ethnic composition of children of immigrants. Additionally, adjusting for family demographics does not change anything. The full model specifications indicate only tiny disadvantages directly associated with children's migration background at ages of 3 and 5 but not further disadvantage at later ages (see Table 6), The same patterns hold when looking separately at verbal and quantitative skills, although the migrant gap in quantitative skills vanish entirely after adjusting for covariates (verbal skills, see Table 7; quantitative skills, see Table 8).

Table 5 Economic, cultural, and demographic composition of migration groups.

	Migration background		
	No immigrant parents (native)	One immigrant parent	Two immigrant parents
Parental education			
Low	53.9	39.7	36.8
Medium	25.1	25.3	26.6
High	21.0	35.0	36.6
HH Income quintile			
1st	7.6	14.6	16.7
2nd	17.5	22.4	37.2
3rd	20.0	16.2	18.1
4th	24.3	16.4	9.7
5th	30.6	30.4	18.3
Ethnicity			
White	94.9	46.6	13.6
Mixed	2.6	7.4	6.8
Indian	0.5	6.9	15.2
Pakistani or Bangladeshi	0.6	21.5	37.9
Black or Black British	1.1	13.9	9.4
Another ethnicity	0.2	3.8	17.0
Language spoken at home			
Only English	97.8	53.8	14.4
English in part	1.9	37.2	57.8
No English	0.3	9.0	27.9
Siblings at home			
None	16.5	14.8	13.0
One	48.9	38.7	31.0
Two or more	34.6	46.5	56.0
Family type			
Two parents	74.8	80.2	96.2
Single mum/dad	19.9	17.7	3.5
Other	5.3	2.1	0.4
Mother's age at birth			
20 or younger	11.5	7.2	6.9
21–25	16.5	20.0	24.2
26 or older	72.0	72.9	68.9
N	7917	1091	501

Notes: All data weighted. Case numbers unweighted.

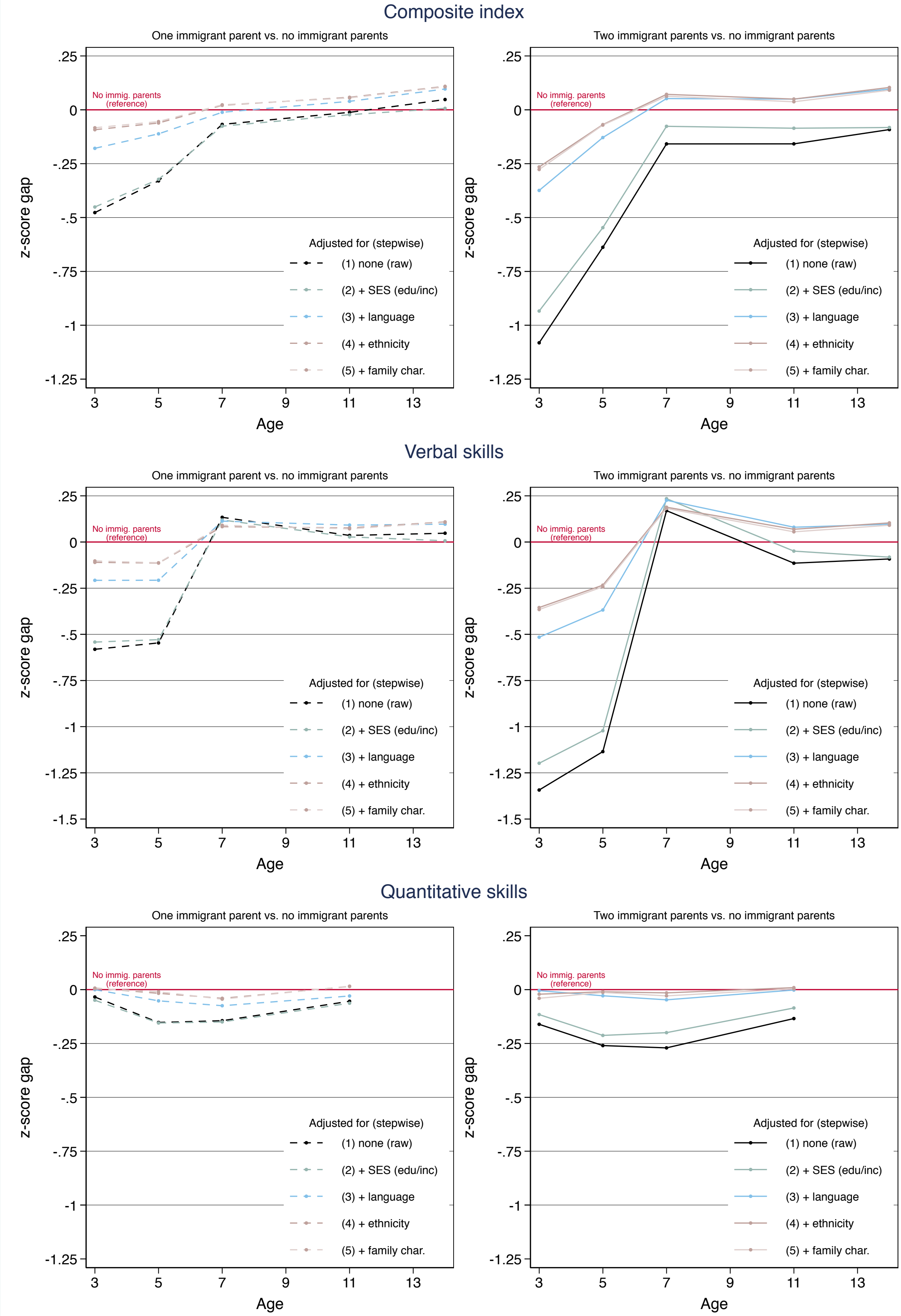


Figure 7 Achievement gaps by migration status step-wisely adjusted.

Notes: 'no immigrant parents' (natives) is the reference (red line).

Table 6 Multivariate models for composite ability index (full model specification).

	Composite index at age …				
	3	5	7	11	14
Education (ref. low)					
medium	0.14*	0.17*	0.18*	0.16*	0.15*
high	0.40*	0.39*	0.46*	0.42*	0.51*
Income (ref. 1st Q)					
2nd Quintile	0.13*	0.15*	0.09	0.05	0.09*
3rd Quintile	0.40*	0.36*	0.33*	0.28*	0.15*
4th Quintile	0.55*	0.47*	0.48*	0.37*	0.20*
5th Quintile	0.66*	0.55*	0.60*	0.53*	0.35*
Migration (ref. native parents)					
1 immigrant parent	–0.08*	–0.05	0.02	0.06	0.11
2 immigrant parents	–0.28*	–0.07	0.06	0.04	0.09
Ethnicity (ref. White)					
Mixed	–0.01	0.01	–0.04	–0.01	0.05
Indian	–0.21*	–0.13	–0.01	0.07	–0.16
Pakistani/Bangladeshi	–0.58*	–0.44*	–0.21*	–0.25*	–0.08
Black	–0.40*	–0.21*	–0.16	–0.05	–0.09
other	–0.29*	–0.12	0.06	0.14	0.06
Language at home (ref. English only)					
English in part	–0.34*	–0.30*	–0.08	–0.05	–0.16*
No English	–0.70*	–0.43*	–0.13	–0.19	–0.22*
Female (ref. male)	0.25*	0.16*	0.10*	0.01	0.03
Siblings at home (ref. none)					
One	–0.18*	–0.08*	–0.05	–0.05	–0.09*
Two or more	–0.37*	–0.18*	–0.14*	–0.10*	–0.16*
Disrupted family (ref. two parents)	–0.11*	–0.04	–0.07*	–0.06	–0.05
Mother's age at birth (ref. < 21)					
21–25	0.08	0.05	0.05	–0.00	0.04
26+	0.15*	0.09	0.07	0.08	0.18*
Constant	–0.44*	–0.47*	–0.52*	–0.42*	–0.38*
R^2	.30	.16	.15	.12	.11

Notes: N= 9509. M=5 imputation datasets. Significance: * p < .05.

Table 7 Multivariate models for verbal skills (full model specification).

	Verbal skills at age …				
	3	5	7	11	14
Education (ref. low)					
medium	0.10*	0.15*	0.15*	0.13*	0.15*
high	0.27*	0.37*	0.43*	0.35*	0.51*
Income (ref. 1st Q)					
2nd Quintile	0.10	0.14*	0.07	0.06	0.09*
3rd Quintile	0.36*	0.33*	0.28*	0.22*	0.15*
4th Quintile	0.48*	0.44*	0.41*	0.32*	0.20*
5th Quintile	0.51*	0.53*	0.51*	0.45*	0.35*
Migration (ref. native parents)					
1 immigrant parent	–0.10*	–0.11*	0.09	0.07	0.11
2 immigrant parents	–0.37*	–0.24*	0.18*	0.06	0.09
Ethnicity (ref. White)					
Mixed	–0.03	–0.02	0.10	0.11	0.05
Indian	–0.37*	–0.13	0.22*	0.14	–0.16
Pakistani/Bangladeshi	–0.70*	–0.60*	0.14	–0.29*	–0.08
Black	–0.39*	–0.43*	0.16	0.21*	–0.09
other	–0.58*	–0.48*	0.03	0.01	0.06
Language at home (ref. English only)					
English in part	–0.40*	–0.45*	–0.06	–0.06	–0.16*
No English	–0.80*	–0.80*	–0.09	–0.10	–0.22*
Female (ref. male)	0.23*	0.05*	0.17*	–0.06*	0.03
Siblings at home (ref. none)					
One	–0.14*	–0.13*	–0.09*	–0.04	–0.09*
Two or more	–0.27*	–0.25*	–0.25*	–0.12*	–0.16*
Disrupted family (ref. two parents)	–0.10*	–0.06*	–0.08*	–0.07*	–0.05
Mother's age at birth (ref. < 21)					
21–25	0.05	0.02	0.00	0.03	0.04
26+	0.11*	0.12*	0.08	0.09	0.18*
Constant	–0.32*	–0.29*	–0.47*	–0.35*	–0.38*
R^2	.27	.24	.13	.10	.11

Notes: N= 9509. M=5 imputations. Statistical significance: * $p < .05$

Table 8 Multivariate models for quantitative skills (full model specification).

	Quantitative skills at age …			
	3	5	7	11
Education (ref. low)				
medium	0.09*	0.14*	0.15*	0.12*
high	0.33*	0.29*	0.37*	0.31*
Income (ref. 1st Q)				
2nd Quintile	0.02	0.09	0.06	0.02
3rd Quintile	0.10*	0.24*	0.25*	0.21*
4th Quintile	0.16*	0.30*	0.36*	0.26*
5th Quintile	0.29*	0.34*	0.49*	0.38*
Migration (ref. native parents)				
1 immigrant parent	0.00	–0.01	–0.04	0.02
2 immigrant parents	–0.04	–0.01	–0.03	0.00
Ethnicity (ref. White)				
Mixed	0.06	–0.07	–0.02	–0.12
Indian	0.15	–0.07	–0.03	–0.04
Pakistani/Bangladeshi	–0.11	–0.22*	–0.36*	–0.11
Black	–0.04	–0.17	–0.11	–0.28*
other	0.13	0.10	–0.00	0.21
Language at home (ref. English only)				
English in part	–0.08	–0.20*	–0.04	–0.02
No English	–0.24*	–0.09	–0.11	–0.20*
Female (ref. male)	0.02	0.20*	–0.01	0.07*
Siblings at home (ref. none)				
One	–0.15*	–0.02	–0.05	–0.04
Two or more	–0.29*	–0.08*	–0.08*	–0.04
Disrupted family (ref. two parents)	–0.11*	–0.04	–0.07*	–0.02
Mother's age at birth (ref. < 20)				
21–25	0.08	0.07	0.03	–0.04
26+	0.15*	0.06	0.03	0.04
Constant	–0.16*	–0.41*	–0.33*	–0.32*
R^2	.08	.07	.10	.07

Notes: N= 9509. M=5 imputations. Statistical significance: * $p < .05$

5.5.3 Explaining primary school gaps by preschool inequalities

How predictive are preschool inequalities for later inequalities in school? How do socio-economic status of the family and migration background shape children's achievement during school years?

In the second part of the overall analysis, we are interested to analyse achievement gaps in school age while taking into account initial gaps right at the time when children started school (age 5). Our aim is therefore to estimate additional effects of socio-economic status and migration background on later achievement which we identify by conditioning on children's achievement at school entry (see Equations 3 and 4 in the method section). This estimation allows us to study (1) how much of the later gaps in school are explained by school entry gaps and, complementary,

(2) how much of the later gaps are explained by the additional influence of covariates after children entered school. In the following, we refer to the additional influence of, for example parental education, as the direct effect of parental education that remains after we took into account the effects of age 5 achievement on later achievement at ages 7, 11, and 14. By comparing the overall gaps at school ages (total effect) with the residual gaps (direct effect) we get an estimate for the part of the total effect that is attributable to earlier inequality mechanisms operating *before and up to school entry age* and their consequences for later achievement (indirect effect) and a part of the total effect that is not attributable to earlier inequality but to inequality mechanisms operating *after children* entered school age (direct effect). We carried out this decomposition into total, direct, and indirect effects first not conditional on the other covariates ('overall' gaps in the following) and, second, conditional on covariates ('net' gaps).

5.5.3.1 Parental education

Estimates for the decomposition into total, direct, and indirect effects (parental education effects operating through earlier performance) not adjusted for other covariates are presented in Figure 8 (separate panels for different outcomes). Covariate adjusted decomposition estimates are presented in Figure 9. All estimates can be found in the Appendix (see Appendix 5.6). Furthermore, Table 9 facilitates the interpretation by informing about the percentage of school-age achievement gaps by parental education that is attributable to earlier achievement differences (age 5) and the additional role that parental education plays beyond preschool.

First of all, both figures (adjusted and unadjusted) reveal that earlier achievement inequalities explain later school-age inequalities to a considerable extent. Take for example findings for the composite ability index: Table 9 shows that almost 70% of the low-high parental education overall (unadjusted) gap at age 7 is explained by the children's differential achievement at age 5. The same figure for the net (adjusted) gap is around 80%. The explanation for that can be sought in the persistency of achievement – children how performed better at earlier stages will on expectation perform better at later stages too. Yet, this persistency is imperfect to the extent that there is heterogeneity in children's learning trajectories causing individual mobility in our relative (z-score) achievement measures: some children are gaining while others are losing ground in relative terms.

The direct effects of parental education provide summary estimates for the social direction of such mobility: children with high educated parents are more likely to gain ground whereas children with low educated parents are more likely to lose ground. Since the direct effects show a tendency to increase over age, we may conclude that this mobility advantages and disadvantages cumulate – in other words, children from higher educated parents accumulate additional advantage over years of schooling. By age 14, only 43% of the overall (unadjusted) low-high gap is explained by age 5 achievement (see Table 9). The time-trend is less clear for verbal skills for which direct effects are particularly large (less than 50% of verbal skill gaps are explained by earlier gaps in verbal skills). In contrast, earlier achievement gaps in quantitative skills are substantially carried over to later school ages. More than 90% of the high-low gap at age 7 and 70% of the gap at age 11 are explained by earlier discrepancies in quantitative skills. Hence, albeit on a smaller level, additional social inequality in quantitative skills accumulate over school life. The main conclusions remain substantially unchanged when we condition the analysis on other covariates (including income, migration background, language at home, ethnicity, child's gender, and family characteristics; see the adjusted gaps in Table 9).

Finally, we were interested to study how the additional effect of parental education beyond the preschool period operate at different points of the distribution of preschool achievement. *Is parental education particularly relevant for later achievement among low preschool performers (shelter/compensation hypothesis)? Or is the additional role of parental education more relevant at the top (boosting hypothesis)?*

To address these questions, we estimated the structural model is described in Equation 5 (but here for groups of parental education, for estimates see Appendix 5.7). Based on the results, we classified children into 'low' (z-score = –1), 'average' (z-score = 0) and 'high' performers (z-score = 1) at age 5. For quantitative skills we used smaller of –/+ .5 SD around the average. Departing from a certain performance level, we simulated potentially diverging achievement trajectories based on the estimates of the interaction model. Figure 9 shows the results.

The additional influence of parental education over school years seems to concentrate first more at the bottom of the distribution of achievement at school start: when starting low, children from higher educated parents seem to fare better over time compared to children from lower educated parents who perform similar at school entry age. However, at age 14 we see compensation at the low performance stratum but exacerbation at the high-performance stratum. Patterns are most pronounced for verbal skills as there seems to be more mobility over school years. In contrast, for quantitative skills we note more persistency and only little additional influence of parental education. Nonetheless, also when concerning quantitative skills, we see that it is the children from higher educated parents who show more upward and less downward mobility than children from lower educated parents.

Table 9 Percentage of age-specific gaps by parental education explained by age 5 achievement.

	Age 7	Age 11	Age 14
Gaps		Composite index	
Overall			
Low vs. high	68	55	43
Medium vs. high	63	49	33
Net			
Low vs. high	79	64	43
Medium vs. high	73	57	34
		Verbal skills	
Overall			
Low vs. high	38	47	32
Medium vs. high	36	44	26
Net			
Low vs. high	58	67	37
Medium vs. high	59	64	31
		Quantitative skills	
Overall			
Low vs. high	93	69	–
Medium vs. high	80	56	–
Net			
Low vs. high	103	73	–
Medium vs. high	91	60	–

Notes: 'Overall' refers to the overall gaps by parental education not controlling for any additional covariates. 'Net' refers to gaps by parental education conditional on covariates of the full models. Note that 100 – percentage yields the percentage of the gaps that is due to the additional effect of parental education on achievement after school entry (age 5).

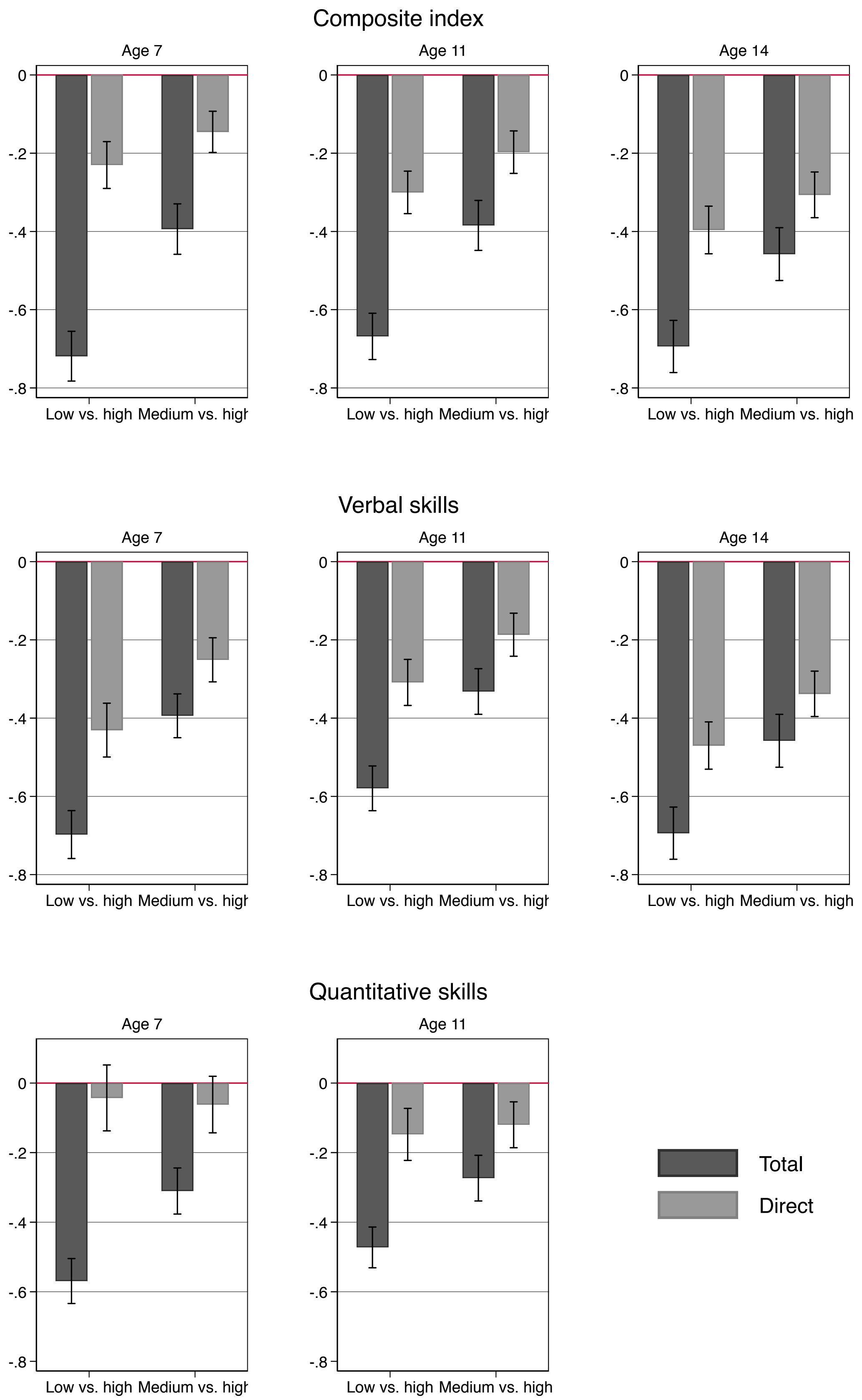


Figure 8 Total and direct effects of parental education (overall/unadjusted).

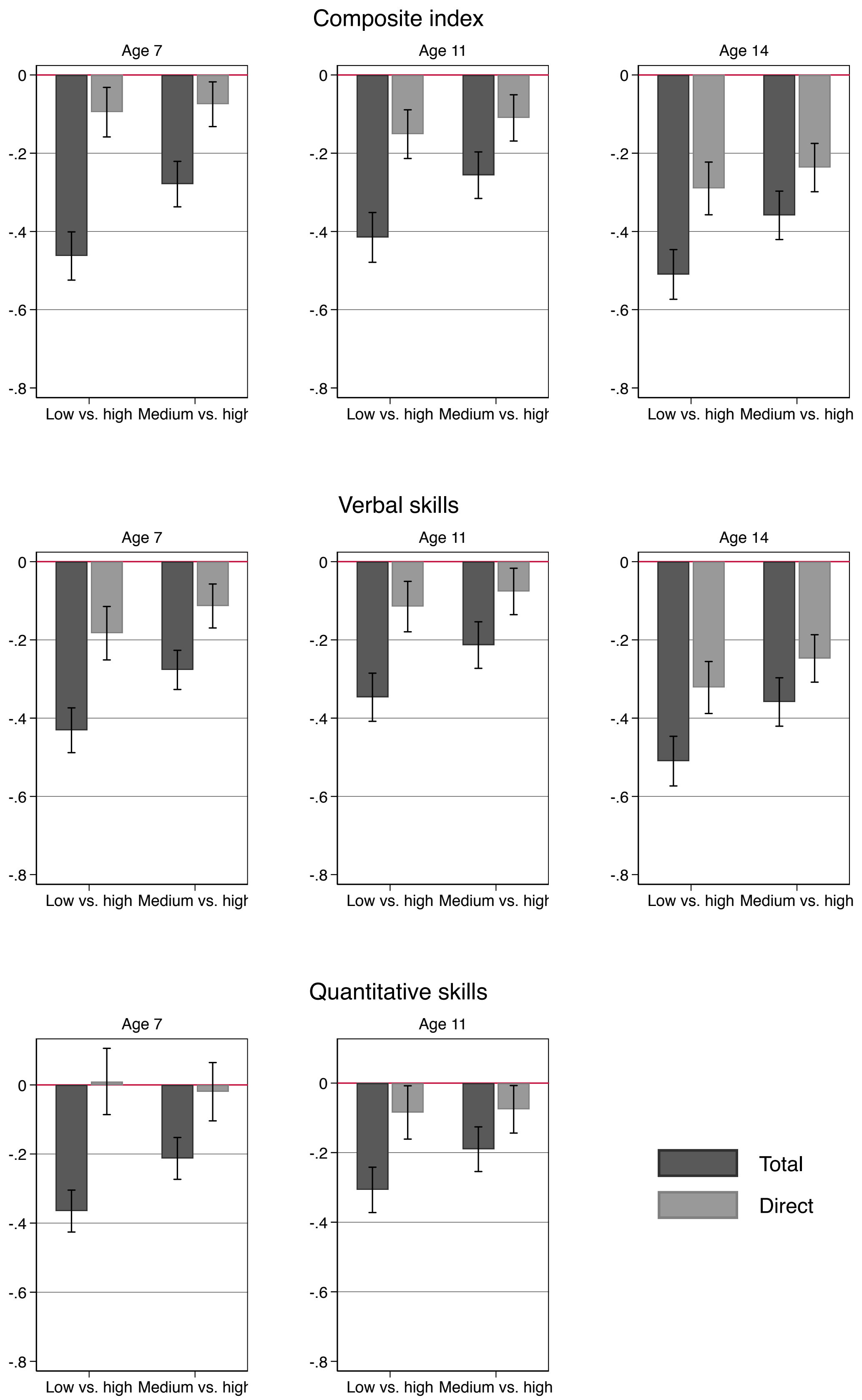


Figure 9 Total and direct effects of parental education (net/adjusted).

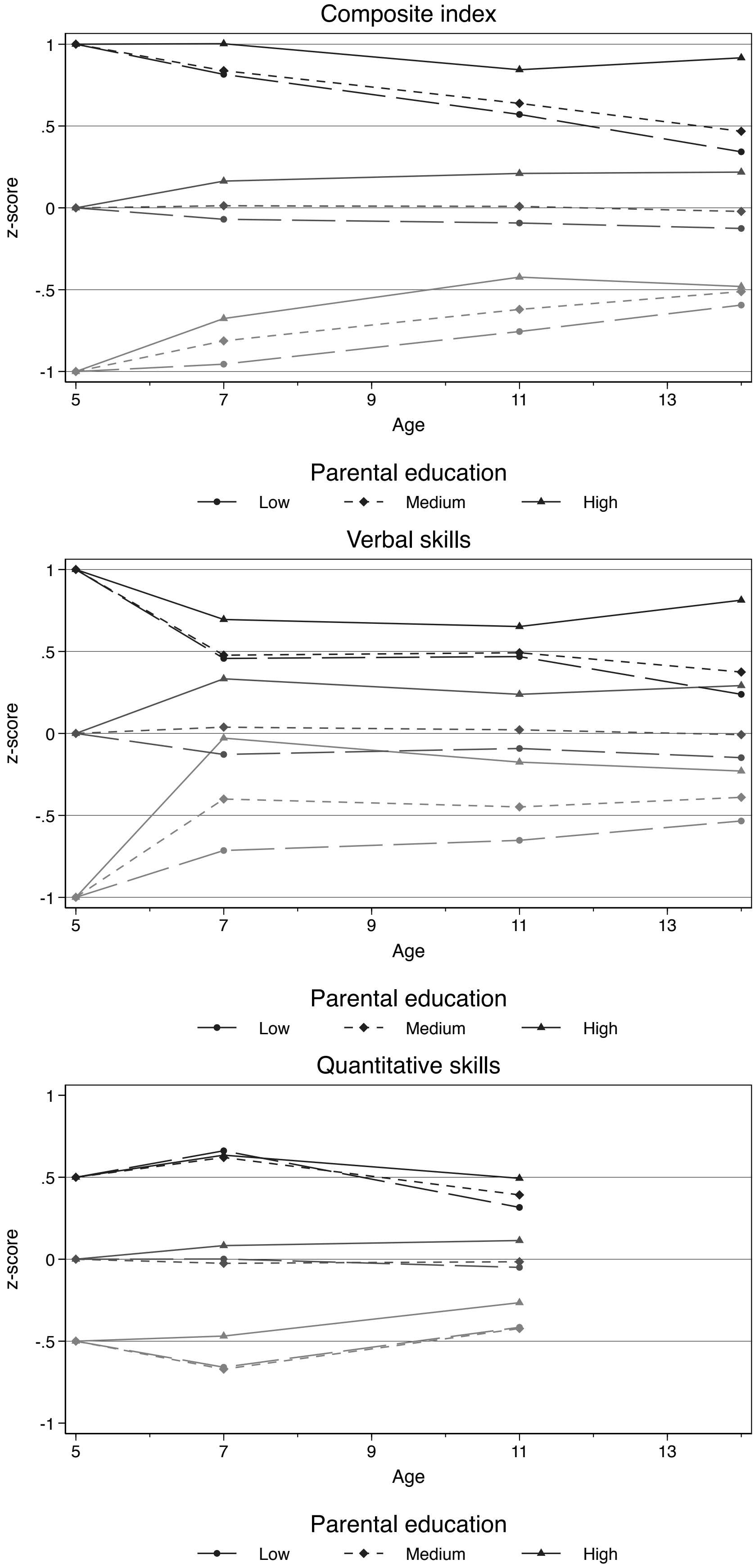


Figure 10 Divergent achievement trajectories by parental education (simulation).

5.5.3.2 Income

In the next step, we replicated the analyses for income-achievement gaps. Figure 11 and Figure 12 show total and direct effects on achievement for income groups overall (not adjusted for additional covariates) and net (adjusted for additional covariates) respectively (see Appendix 5.8 for detailed estimates). Table 10 presents the percentage of income-achievement gaps at schooling ages 7 to 14 explained by earlier performance gaps at school entry age.

Both overall and net gaps by income-class shrink from age 7 to age 14 and that holds true for all outcomes. Furthermore, we see that, to a large extent, income-achievement gaps in school age are explained by income-achievement inequality between children before or right at the start of school (for example, almost 70% of the overall composite ability gap at age 11 and 14). Like for parental education, HH income has the largest additional effect over school for verbal skills (least explained by school entry performance). Nonetheless, remaining direct effects still suggest that over schooling higher income households is associated with additional achievement advantages.

Figure 13 shows divergent trajectories for income-groups (for regression models see Appendix 5.9). We simplified the analysing by collapsing the two lower quintiles to a ‘low income’ group and the upper two quintiles to a ‘high income’ group. The middle quintile group represents a ‘medium income’ group. The results clearly indicate that advantages and disadvantages by parental income play out particularly at the top of the performance distribution at age 5. For instance, among high performers (initial 1 SD above the mean), students from lower income groups consistently lose ground compared to students from upper income groups. Among average and low performers, trajectories for different income groups are comparably similar. The findings are by and large consistent over all outcomes under study. Taken together it seems that high performing children from lower income groups are particularly vulnerable in their relative achievement development: they might start high, but it becomes increasingly difficult for them to keep up with their higher-income class peers. Conversely, a higher income of the family is not associated with much of an advantage among low achievers.

Table 10 Percentage of age-specific gaps by HH income (quintiles) explained by age 5 achievement.

Gaps		Age 7	Age 11	Age 14
			Composite index	
Overall				
	Q1 vs. Q5	87	69	69
	Q3 vs. Q5	69	53	46
Net				
	Q1 vs. Q5	88	72	91
	Q3 vs. Q5	66	51	55
			Verbal skills	
Overall				
	Q1 vs. Q5	64	70	60
	Q3 vs. Q5	45	51	39
Net				
	Q1 vs. Q5	61	72	77
	Q3 vs. Q5	47	53	51
			Quantitative skills	
Overall				
	Q1 vs. Q5	97	78	–
	Q3 vs. Q5	76	61	–
Net				
	Q1 vs. Q5	89	69	–
	Q3 vs. Q5	54	47	–

Notes: 'Overall' refers to the overall gaps by income quintile not controlling for any additional covariates. 'Net' refers to gaps by income quintiles conditional on covariates of the full models. Note that 100 – percentage yields the percentage of the gaps that is due to the additional effect of income on achievement after school entry (age 5).

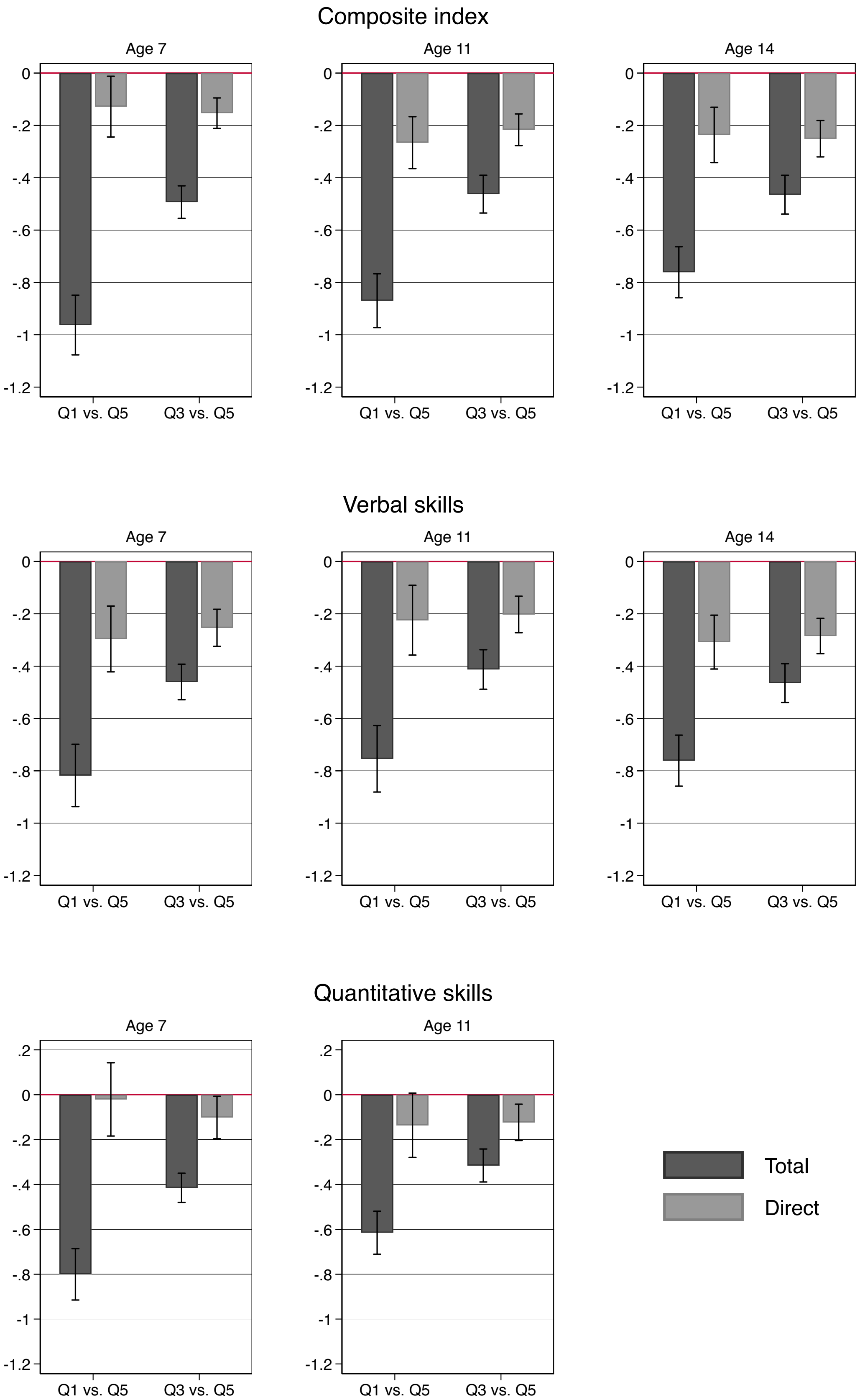


Figure 11 Total and direct effects of HH income (Q1/Q3 vs. Q5, overall/unadjusted).

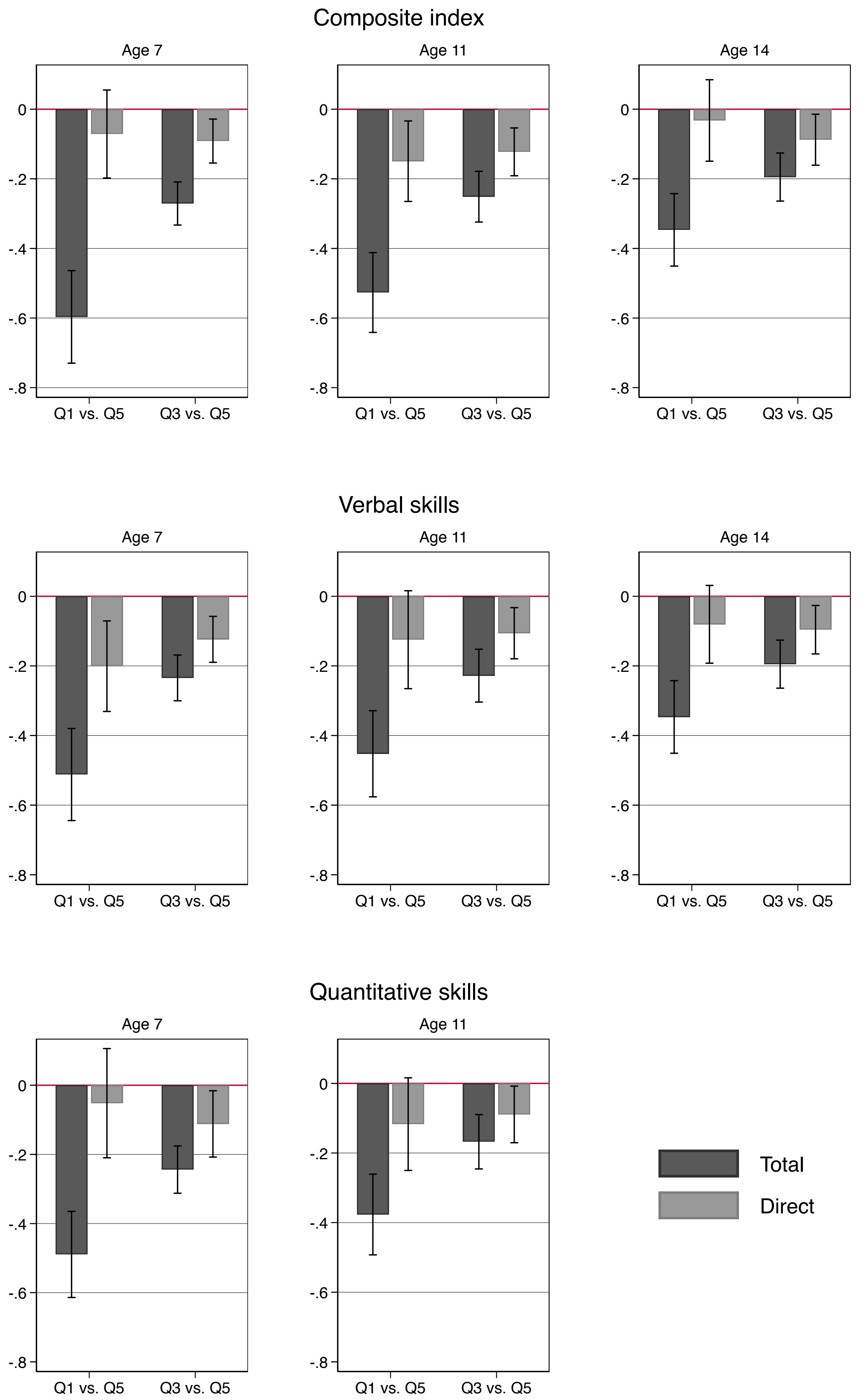


Figure 12 Total and direct effects of HH income (Q1/Q3 vs. Q5, net/adjusted).

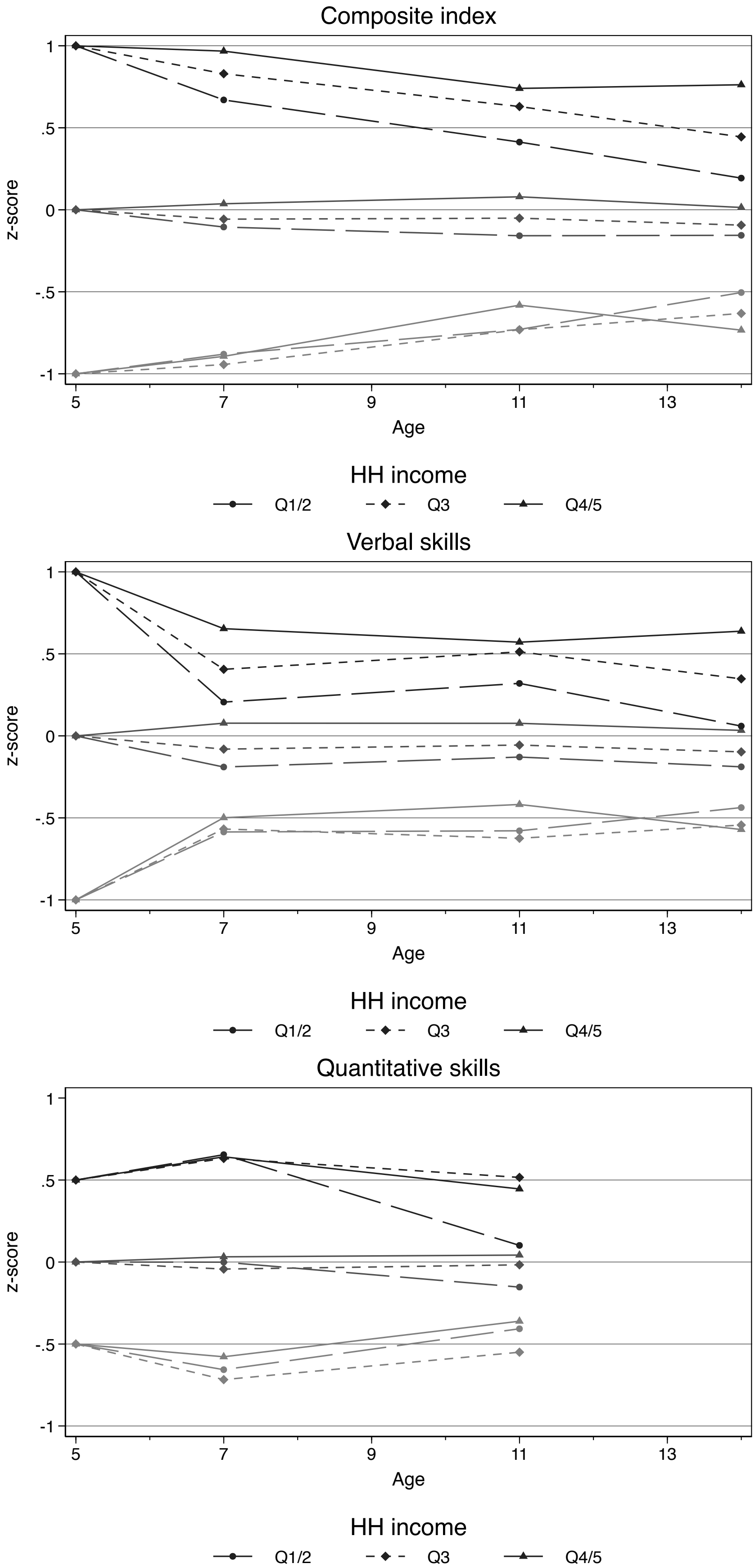


Figure 13 Divergent achievement trajectories by HH income (simulation).

5.5.3.3 Migration background

Figure 14 and Figure 15 show total and direct effects of migration background (overall and net respectively, see Appendix 5.10 for detailed estimates). Compared to socio-economic status, findings for migration background yield a quite different picture. In the analyses on migration-related gaps we have already seen little to no disadvantage for children of immigrants in school age. However, when accounting for earlier performance, we now see that children of immigrants are actually over-performing. Looking for example at the composite index (see Figure 14), we observe actually a premium for children of immigrants over children of natives, particularly for those with two immigrant parents. That finding can be explained by a rapid catching-up effect of migrant children in school. If they had performed at the same level as native students at school entry, they would actually perform better later (however, not significant for quantitative skills). At least for verbal skills, the migrant premium remains significant even when controlling for all other covariates (Figure 15).

The divergent trajectory models (Figure 16) suggest that the migrant premium is particularly pronounced among low performers (for regression models see Appendix 5.11). Take for instance the findings on verbal skills (middle panel in the figure). Among those who enter school at a low performance level (–1 SD), it is the children of immigrants who catch up quickly and reach (and partly surpass) the level of the medium performance native kids by age 7.

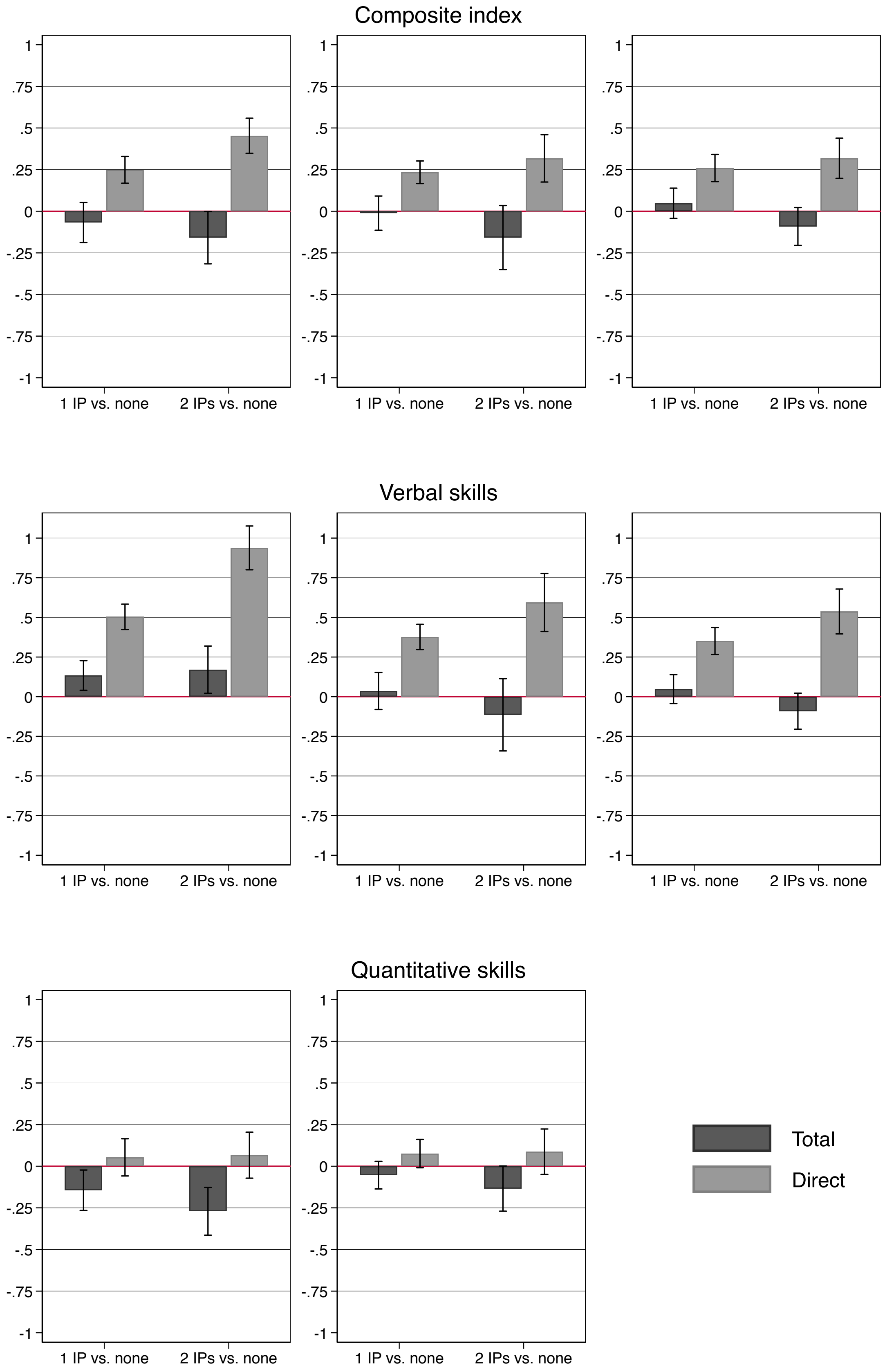


Figure 14 Total and direct effects of migration background (overall/unadjusted).

Notes: 1 IP = One immigrant parent. 2 IPs = Two immigrant parents. None = no immigrant parents.

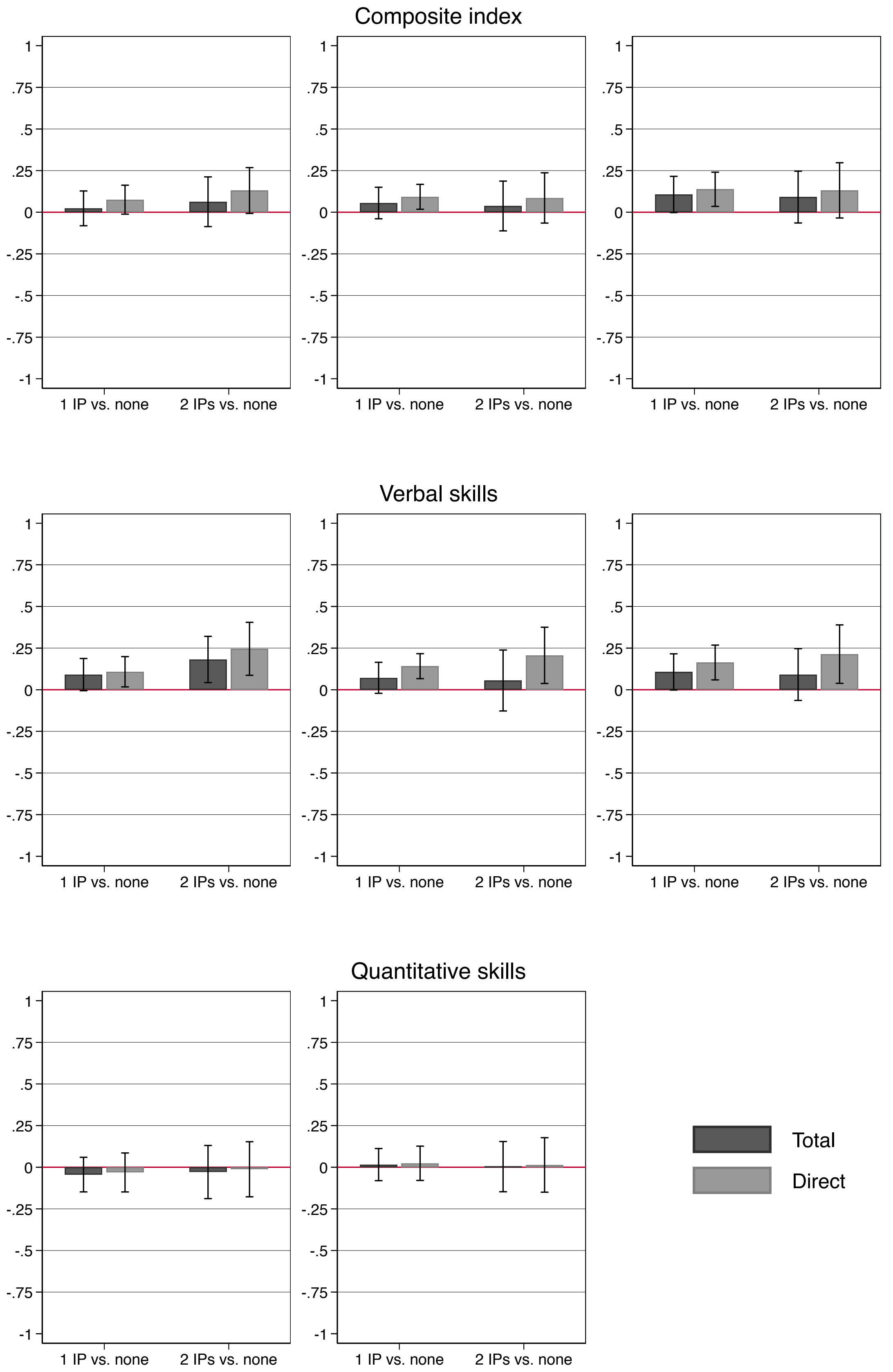


Figure 15 Total and direct effects of migration background (net/adjusted).

Notes: 1 IP = One immigrant parent. 2 IPs = Two immigrant parents. None = no immigrant parents.

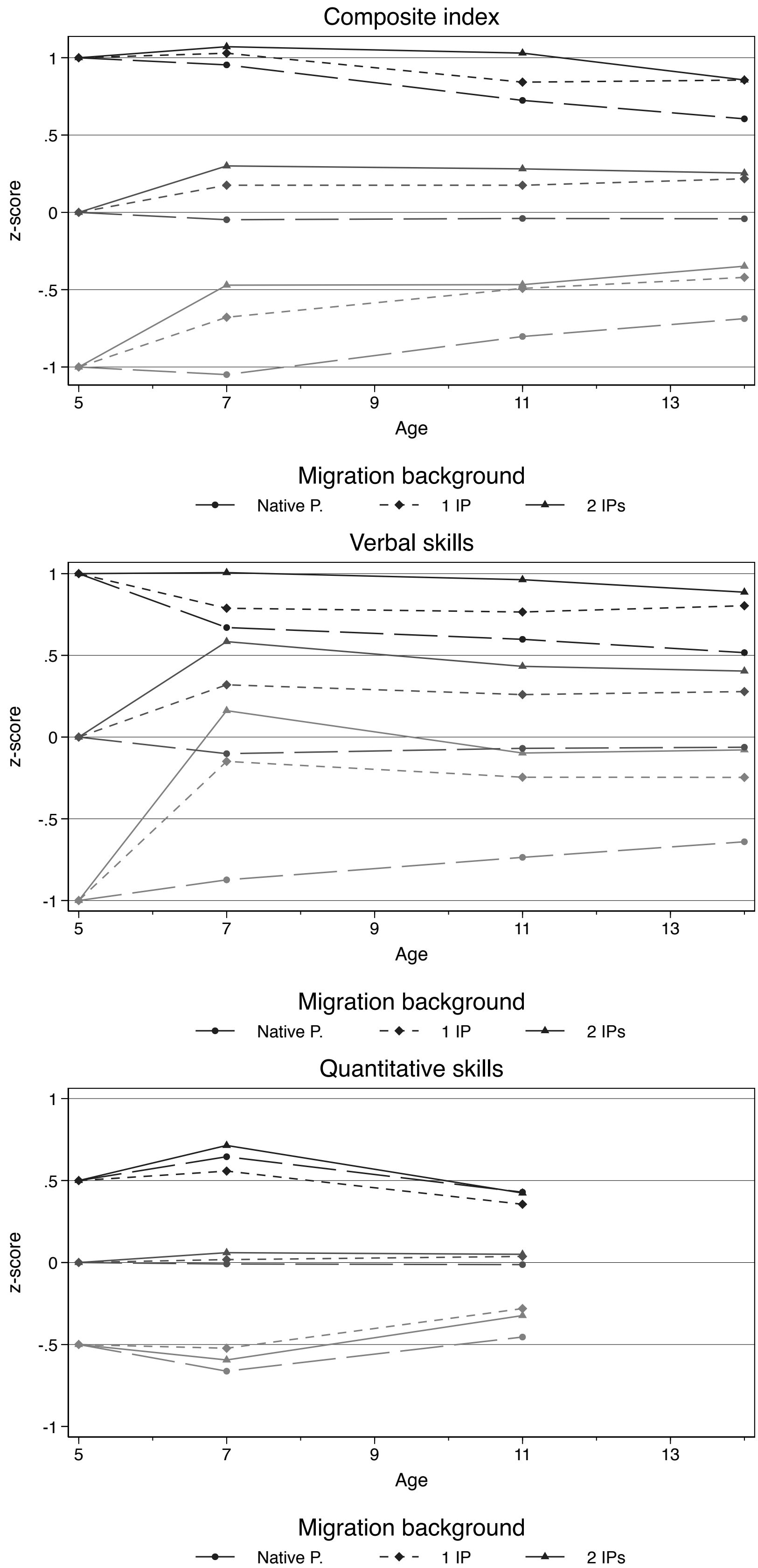


Figure 16 Divergent achievement trajectories by migration background (simulation).

Notes: 1 IP = One immigrant parent. 2 IPs = Two immigrant parents.

5.5.3.4 Ethnicity

Finally, we also inspected direct and total effects of ethnicity in the same fashion as above. We stay here with the aggregate picture and do not present a simulation here as the case numbers for some of the ethnicity groups are too small. Yet, both Figure 17 and Figure 18 (find detailed estimates in Appendix 5.12) suggest premiums for some of the ethnic minority groups over children of white majority status. At the example of verbal skills for the overall effects (Figure 17), we see that when having the same performance level at age 5, the average non-white child would perform better than the average white child at age 7 (highest premium for Pakistani and Bangladeshi and Indian), age 11 (highest premium for black kids), and age 14 (highest premium for 'other' group). Some of these minority premiums remain even when controlling for all other covariates (Figure 18).

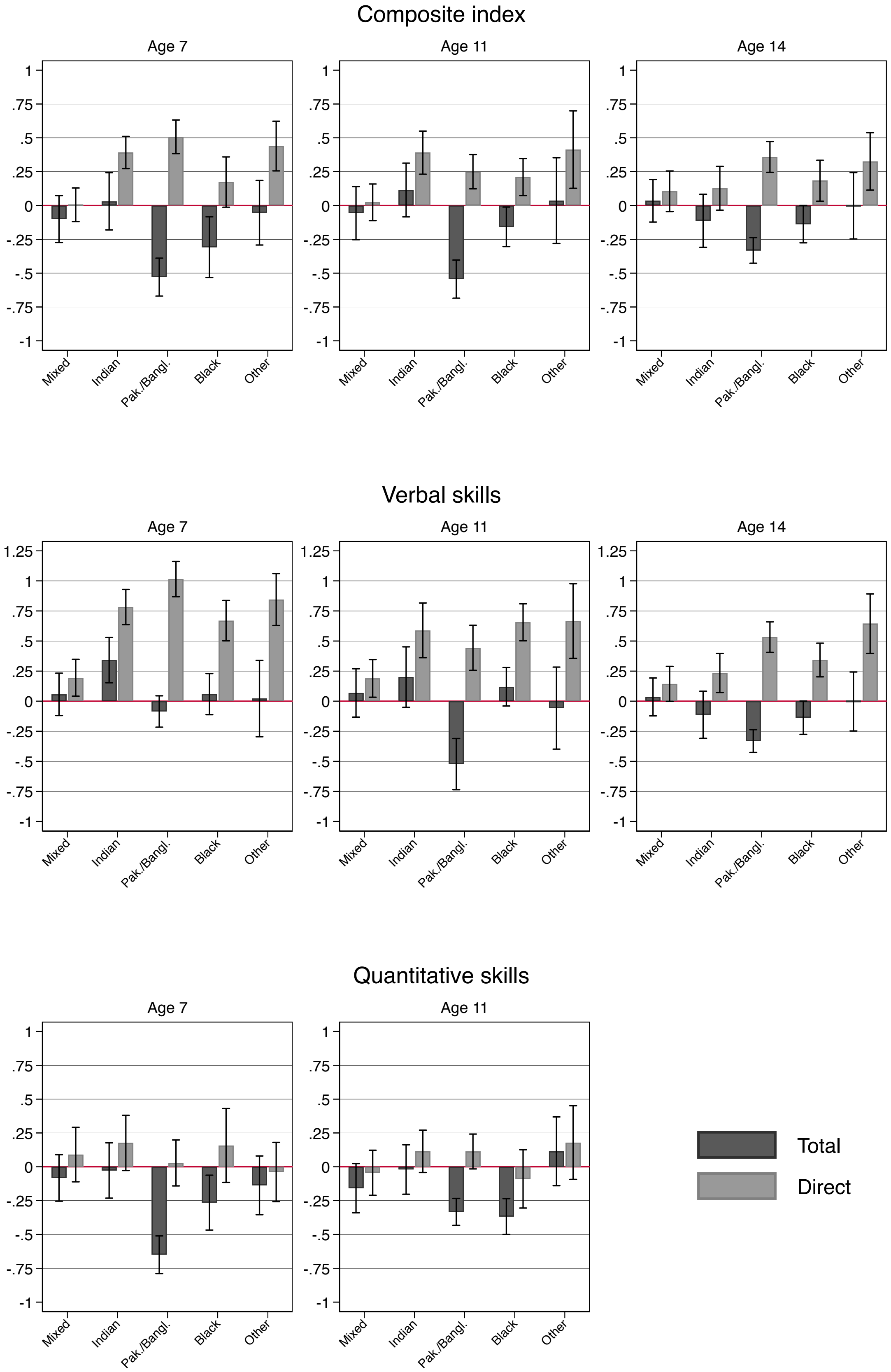


Figure 17 Total and direct effects of ethnicity (overall/unadjusted).

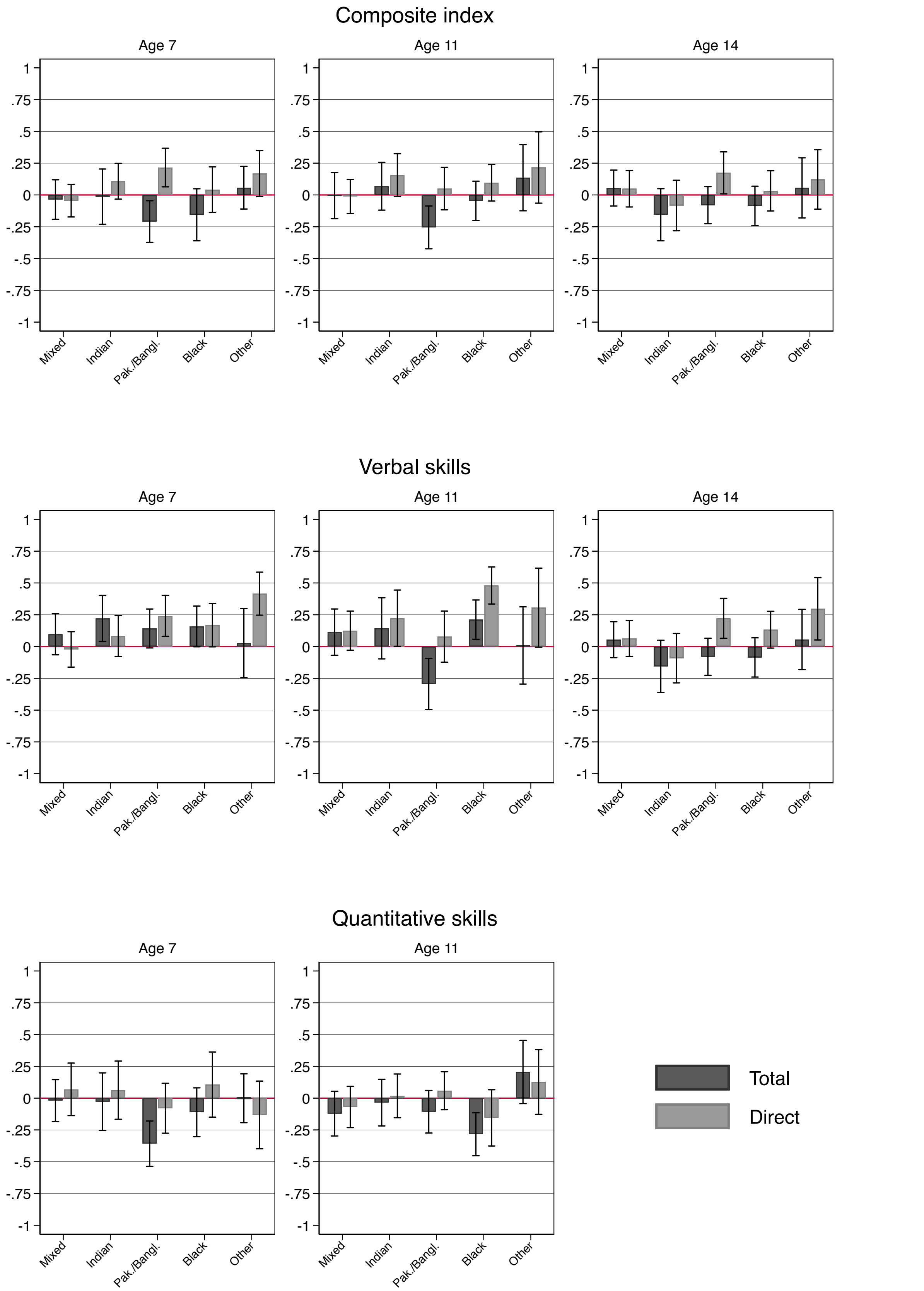


Figure 18 Total and direct effects of ethnicity (net/adjusted).

5.6 Summary and conclusions

Using rich and most recent longitudinal data from the Millennium Cohort Study, this chapter analysed roots and development of social and migration/ethnic-related achievement gaps for children born in 2000–2001 in the UK. Detailed data on cognitive achievement and teacher evaluations allowed us to study achievement gaps over an extended period of the early life course spanning preschool (age 3), the transition to school (age 5), as well as developments during children's primary (age 5 to 11) and early secondary school life (age 11 to 14). The data allowed us to provide differentiated analyses using a composite index measuring students relative positions in several achievement domains, a relative measure for verbal skills related to vocabulary, language, and reading, as well as a relative measure for quantitative skills related to competencies in problem solving, arithmetic, and math.

Overall, we found strong and rather persistent gaps for dimensions of socio-economic status. Most persistent were discrepancies in average achievement by parental education. Income-inequality in achievement plays an important role even though disadvantage by parental income tends to shrink over the observed age span. We also found that SES gaps in achievement that accumulated before children enter school explain a considerable part of SES gaps in school life although there were differences by domains (higher gap persistency for quantitative skills, lower persistency for verbal skills). However, our analyses revealed also that dimensions of parental SES (income and education) continue to shape differential achievement of children during school over and above their cumulative influence up to school start.

Moreover, our analyses provided a detailed picture on the educational experiences of children from minority groups providing diverse angles in terms migration, ethnicity, and cultural subgroups. We found clear evidence for a strong disadvantage in preschool achievement for children of immigrants, children of minority ethnicity status, and children from minority cultural/religious groups face. In addition, our analyses detected important interactions between ethnicity and migration status: it is the children of non-white immigrants who face an especially pronounced penalty in preschool achievement. Furthermore, decomposition analyses have shown SES differences between migrant and non-migrant families explain little of those gaps. Although, on the one hand, migrant families on average are poorer in income they have frequently more education than native families on the other hand. However, language culture at home was found to be a powerful mechanism explaining early disadvantages of migrant kids. Clearly, fewer opportunities at home to attain English proficiency at home puts many migrant children at a considerable disadvantage in the early years. However, many of these disadvantages vanish entirely once children enter the formal school system. As our analyses suggested, there is an astounding catching-up effect – children from most minority groups over-proportionally gain in achievement compared to children of majority status. Controlling for a variety of factors including earlier achievement, we could detect an achievement 'premium' for several minority groups.

Although we could not definitely test all the ample mechanisms that are likely to interplay in bringing out these results, one may speculate that the preschool and school system in UK is institutionally particularly well-prepared in alleviating earlier disadvantages children of minority groups face at home. Finally, we may add that our analyses were limited to children of immigrants and – by the design of the MCS that sampled children born in the UK – was unable to investigate children who were born abroad.

6 ITALY

Social and Migration-related Inequality in Achievement in Primary and Secondary Education

Piergiorgio Lovaglio♠, Stefano Verzillo[1]♣♠, & Giorgio Vittadini♠
♠ Department of Statistics and Quantitative Methods, University of Milan-Bicocca
♣ European Commission, Joint Research Centre (JRC)

6.1 Introduction

This chapter will analyse the evolution of achievement gaps in Italy, as students move from primary to secondary schooling. Our chapter will thereby primarily address research question one, particularly the aspect of how social and migration-related gaps in educational achievement develop over time. Tackling these questions, we make use of population-level data from Italy capturing achievement of all students in Italian primary and secondary schools in a specific year by INVALSI – the Italian National Institute for the Evaluation of the School System. Combining such population-level data with a pseudo-panel design capturing end of primary schooling, lower and upper secondary schooling, our study can reveal with high precision how social and migration-related inequality in achievement evolves in the school career of students. Our study, however, was limited as to the second research question, since at the time of writing this chapter, population data could not be linked across years on the individual level.

Our chapter is structured as follows. First, by providing an overview of the education system in Italy, we locate our case in the overall framework of the report. Second, we describe the population data and discuss our analytical approach implementing a pseudo-panel design. Moreover, we discuss the construction of our central dependent and independent variables. Afterwards, we present our findings on social and migration-related achievement gaps. Our analysis will also inspect the intersection of gender and migration status in patterns of inequality of educational achievement. Finally, we draw a conclusion.

6.2 The education system in Italy

In Italy, children can go to pre-primary school ("Kindergarten") which lasts for three years between the age of 3 and the age of 5. Although attendance is optional, a vast majority of children participate in it. For instance, in 2016 around 95% of Italian children (aged between 3 and 5) attended pre-primary school compared to an average attendance rate of 88% at EU–22 level (OECD 2018c, Education at a Glance Figure B2.1a and 1b). Children enter school by September in the year a child turns to six. The first cycle, primary schooling, consists of 5 school years across

[1] The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of this publication. The preparatory research work for this chapter was undertaken at the University of Milan-Bicocca while estimation of the empirical models and the final drafting of the text have been conducted when Stefano Verzillo took service at the European Commission, Joint Research Centre, Competence Centre on Microeconomic Evaluation (CC-ME).

the typical age range of 6 to 10. Primary schooling is followed by three years of lower secondary schooling from age 11 to 14. After that, students enter the upper secondary cycle lasts for the following five years between the age of 14 and 18, but students can leave school by the age of 16 (compulsory school age).

In general, Italy adopts a mixed tracking model in secondary schooling (Blossfeld et al. 2016). While primary and lower secondary in Italy is rather comprehensive, upper secondary education is formally differentiated by tracks (Contini & Triventi 2016). When entering upper secondary schooling students can decide to attend one of three separate tracks that provide quite distinct educational curricula and are embodied in different school types. The most prestigious and demanding track is the academic track (*Lyceum*). It aims at preparing students for educational maturity to enrol later at higher tertiary education. Hence, *Lycei* as the main route to university focus on teaching broader and rather generalist educational contents closely related to academic competencies. Within the academic track there is some differentiation across the Lyceum schools regarding specialisation (e.g., specialised in science or languages). Alternatively, students can enrol to a technical track (*instituti tecnici*) or vocational track (*instituti professionali*). Both of those tracks provide more practical education linked to non-academic, technical professions. However, the technical track prepares for upper-level jobs and the vocational track for lower level technical jobs. In contrast to other tracking systems like in Germany or the Netherlands, students are not formally allocated to tracks based on their prior educational achievement but instead are free to choose even though lower secondary teachers may issue track recommendations (Contini & Triventi 2016).

Having concluded a five-year upper secondary diploma, an Italian student may choose to attend a higher education bachelor's degree of 3 years followed by two years of master courses. On top, PhD programs represent the highest qualification achievable (with usually a length of 3 years).

Till today, the Italian educational system is characterised by comparably low educational attainment. Recent data from the OECD (2018d) shows that in 2017, 39.1% of Italians aged 25–65 did not attain upper secondary education compared to the OECD average of 20.7%. Even though Italians meet with 42.2% upper secondary level attainment the OECD average versus (42.8%), they lag dramatically behind in attaining tertiary education (18.7% compared to OECD average of 36.9%). To be sure, educational attainment has risen across recent birth cohorts, but the pace of catching up is slow, especially for tertiary education. Therefore, Italy may reach the average of OECD countries only on the long-run.

Against this backdrop, it is important to realise that the Italian educational system is genuinely a public system. Only 5% of students attend private schools.[2] Among the OECD countries, Italy is still one of those countries with lower public (private) expenditure in the educational system in terms of GDP. Primary to post-secondary public (private) spending measured as percentage of GDP in fact was 2.78% (.16%) in 2016 while .54% (.33%) was spent on tertiary education. While expenditure for primary and secondary is close to the OECD average

[2] When parental education and SES are accounted for, the private sector reveals to be on average of lower-quality according to students' performances measured via standardised test scores. Possibly, this is mainly due to the presence of a relative majority of remedial schools within the private sector. In fact, recent research found that private schooling in Italy features a certain degree of heterogeneity (Checchi & Verzillo 2018). Concerning PISA results student achievement's levels of confessional private schools seems to be statistically indistinguishable from public ones while the ones of students attending remedial private schools are statistically worse than those of students attending public schools (Checchi & Verzillo 2018).

tertiary expenditure is dramatically lower.

In addition to this, several studies have pointed out how Italy is one of the countries with low intergenerational mobility – with "*less intergenerational upward mobility between occupations and between education levels*" (Checchi et al. 1999). As a result, students' educational and employment careers are shaped by their parents' education to a considerable extent (Checchi 2003; Bratti, Checchi & Filippin 2007). For instance, compared to other educational systems with tracking such as Germany, social inequality (by parental education) in track allocation is remarkably substantial in Italy which has been explained by a less efficient sorting based on abilities in the Italian case (Checchi & Flabbi 2006). Finally, a relevant divide in educational attainment between students from the south and north is still a relevant issue in the Italian system (Bratti et al. 2007).[3]

6.3 Data and methods

Our chapter will analyse achievement gaps in Italy using population data collected by the Italian National Institute for the Evaluation of the School System (INVALSI)[4]. On an annual basis, INVALSI carries out standardised tests to assess students' proficiency levels at various grades. For our purposes, we exploited test data on literacy and math. Currently, the INVALSI database covers the years 2009 to 2016 with measurements taken in each year in four grade levels: Grade 2 (7-year-olds), Grade 5 (10-year-olds), Grade 8 (13-year-olds) and Grade 10 (15-year-olds). Unfortunately, at the present stage, it is still not possible[5] to link individual student data over time since student identifiers change at each measurement occasion. That circumstance precludes at this time a credible longitudinal data analysis based on the individual level. Instead, we treat the data as repeated cross-sectional data at different ages.

However, in the context of this report, a significant advantage of INVALSI data is the fact that achievement tests had been collected for the *whole population* of students in each grade since test participation is mandatory in Italy. Compared to other survey datasets used in this report (such as the NEPS analysed in Chapter 2 or the MCS data in Chapter 5), population data are not plagued by issues of sample selection or attrition. Hence, INVALSI data allows us to achieve a maximum of accuracy in studying the distribution of students' achievement at each grade in a particular year and social and migration-related gaps therein for the whole population of Italian students being in that grade in that year.

6.3.1 Pseudo-panel design

Despite the cross-sectional nature of the data we aimed at mimicking as much as possible a true longitudinal design similar to the other country analyses in this report. We did so by focussing on a specific cohort at grade j in a given year t and then shifting the observed window by k years. By that, we constructed a 'quasi-follow-up' of students belonging to the same cohort at time $t+k$ in grade $j+1$. Essentially, this strategy follows the same logic as the accelerated

---

[3] Bratti and colleagues (2007) provide evidence of how the North-South divide in PISA scores seems to be mainly due to differences in the endowments of students, the efficiency of schools and the socio-economic environment.

[4] Source: INVALSI – estimate of the authors on INVALSI data – INVALSI Statistical Office

[5] A pilot program to build an individual panel dataset following students over different grades has been recently started at the INVALSI Institute; however, it is not yet completed at the moment of writing.

longitudinal design applied in Chapter 2 and 3 for the German and the Dutch case respectively, however with just one time point for a set of students. A natural limitation of our approach is that there is no way of accounting for school dropout or transfers which should be kept in mind when interpreting our results.

Table 1 illustrates our pseudo-panel approach for following up a cohort of five-graders in the school year 2010–2011. The first data point consists of the students being in Grade 5 in 2010–11. This cohort is measured again in Grade 8 using the cross-section from 2013–2014. Finally, the cohort is measured again in Grade 10 by the cross-section in 2015–2016. Hence, our design allows observing the development of achievement gaps as a cohort of more than 500,000 students who have been born in 2000 (the regular students) progress through primary and secondary schooling. Our study design is in line with an earlier ISOTIS report on achievement gaps utilising cross-national assessment data for a pseudo-panel approach (Rözer & Van De Werfhorst 2017) and similar previous research (Dämmrich & Triventi 2018).

We start with 526,462 students at Grade 5. Three years later at Grade 8 the sample size for the cohort slightly reduces to 520,917 students (–1.05%) while subsequently a significant reduction of around 26% is recorded at the second year of the upper secondary school (Grade 10). The substantial sample reduction at Grade 10 can be explained by student dropout which is likely to be selective by educational achievement and should be considered when interpreting the findings.

Table 1 Pseudo-panel design using available INVALSI data.

	Grade, Age and Stage		
School year	Grade 5 Age 10 *Primary (V)*	Grade 8 Age 13 *Lower secondary (III)*	Grade 10 Age 15 Upper secondary (II)
2010–2011	**X** N=526,462	–	–
2011–2012	–	–	–
2012–2013	–	–	–
2013–2014	–	**X** N=520,917	–
2014–2015	–	–	–
2015–2016	–	–	**X** N=383,255

Notes: **X** demarks cross-sectional data used for the analysis.

6.3.2 Main variables

6.3.2.1 Outcomes

Our outcome variables are student test scores measuring educational achievement in Literacy and Math. Even if good psychometric measures (IRT–Rasch scores) are directly provided by the INVALSI institute to ease the interpretation, in order to ensure comparability to the other countries (which adopt and collect data differently one from each other) we implement a two-step strategy to standardise original test scores. First, we derived the normalised ranked position of children (normalised ranks range from 0 to 100) in the achievement distribution separately for each wave. Second, we z-standardized the normalised ranks (z-scores). That is a suitable transformation of the original data since an individual z-score reflects the relative position of a student in the overall distribution but in a continuous metric, also providing correct standard errors, confidence intervals and ensuring the validity of inference of the obtained results. Notice that when outcome variables Y are normally distributed (as strongly supported by data in our case for both outcomes in each wave) this two steps strategy coincides *de facto* with the classic standardization process ($z_Y = (Y-\mu_Y)/\sigma_Y$)) of an outcome variable Y, since ranks are empirical quantiles of a normal cumulative distribution function (for details see theory on fitting distribution with qq-plots, see Shapiro and Wilk 1965).

6.3.2.2 Covariates

In addition to students' educational achievement, INVALSI collected background information from student and parent questionnaires. The main covariates we use is the highest *parental education* as a measure for socio-economic status of the student's family origin and *migration background*.

For *parental education*, we applied the dominance criteria by taking the highest educational level among a student's parents at the moment of the survey participants. We classified parental education into low (12 years of education or less), medium (13 to 15 years of education), and high (more than 15 years of education). Unfortunately, different to other educational surveys (such as PISA[6]), the INVALSI survey does not provide other relevant information related to the socio-economic dimension such as parental income, home ownership, familial educational resources or similar indicators.

For capturing *migration background* of Italian students, we used a threefold classification: Native (no migration background); first-generation migrant (child born abroad); and second-generation migrant (child born in Italy, but at least one parent born abroad). Unfortunately, the scientific use file of the INVALSI data does not allow distinguishing between specific groups of migrant children regarding ethnicity or country of origin.

6.3.3 Data description

Table 2 presents descriptive statistics on the distribution of covariates and outcome variables. Parental education is missing for around 17% and 20% of students of Grade 5 and 8 respectively while only 10% for students in Grade 10. The distribution of parental education is reasonably constant between Grade 5 and 8 (14% high, 35% medium and 30% low) while it changes

6 The OECD Programme of International Student Assessment (PISA) provides a composite index of economic, social and cultural status based on parents' education, family wealth, cultural possession at home, ICT resources, home possession and home educational resources. See Vol II of the PISA Results for more details.

considerably at Grade 10 with a lower proportion of students from disadvantaged families (from 30% to 22%) and larger fractions of students from medium (from 35% to 42%) and high (14% to 24%) parental education status. The reduction of students from disadvantaged socio-economic backgrounds in Grade 10 in favour of medium and high SES when compared with earlier grades clearly indicates the existence of significant differences in drop-out probabilities by SES in Italy (O'Higgins et al. 2007, Checchi 2010).

Students' migration background shows a stable distribution of natives and first- and second-generation migrant students among the different grades. A vast majority of natives (around 87%) cohabits in the school system with around 5% of both first- and second-generation students in the primary school as well as in lower and upper secondary levels. Significantly different between the primary and secondary school grades is the percentage of student non-reporting information on their ethnicity. The 5% of missing data in Grade 5 to 1% reduces to .8% in Grade 8 and 2.4% in Grade 10.

Finally, Table 2 reports simple descriptive statistics regarding the standardised achievement scores in Literacy and Math. As expected, z-score measures for math and literacy were highly correlated with the corresponding original test scores (corr > .97 for Math, corr > .96 for Literacy in all the considered grades).

Table 2 Distribution of parental education, migration background and z-scores by grade.

	Grade 5	Grade 8	Grade 10
Parental education (column %)			
Missing	17.6	20.6	10.2
Low	31.5	30.0	22.8
Medium	36.2	35.0	42.2
High	14.7	14.4	24.8
Total	100.0	100.0	100.0
Migration background (column %)			
Missing	5.2	0.8	2.4
Native	86.0	89.3	87.6
1st generation migrant	4.4	5.2	4.8
2nd generation migrant	4.4	4.7	5.3
Total	100.0	100.0	100.0
Total number of students	526,462	520,917	383,255
Educational achievement (z-scores)			
Literacy			
Mean	0	0	0
Standard deviation	1	1	1
Min / Max	–3.18 / 2.77	–4.21 / 3.89	–2.67 / 3.34
Valid cases	515,104	520,917	371,882
Math			
Mean	0	0	0
Standard deviation	1	1	1
Min / Max	–3.27 / 3.18	–4.72 / 4.32	–2.57 / 3.57
Valid cases	508,615	520,917	383,255

6.4 Results

6.4.1 Achievement gaps by parental education

We start by inspecting overall patterns of inequality in math and literacy achievement by parental education. Least-square means for groups were calculated after accounting for age. Figure 1 shows how average z-score levels develop across Grade levels for children from different social backgrounds concerning parental education. Additionally, Figure 2 plots for each outcome variable and each grade the gaps in math and literacy achievement using high parental education as the reference category. All point estimates for these and following results are available in the Appendix of this report (Appendix Section 6).

Results for both domains, math and literacy, indicate strong associations of achievement and socio-economic background of children. Figure 2 suggests that social inequality in achievement, particularly in terms of the gaps between children from low and high educated parents, are most pronounced for the literacy domain. Moreover, those SES gaps in achievement appear to be quite stable over subsequent grades. All gaps between students from low, medium and high educated parents are statistically significant. Although our observation could not capture earlier developments in gaps, it is likely that those gaps resemble gaps in cognitive achievement that arise very early in children's educational careers (Fernald et al. 2013; Lee & Burkam 2002; Magnuson et al. 2004).

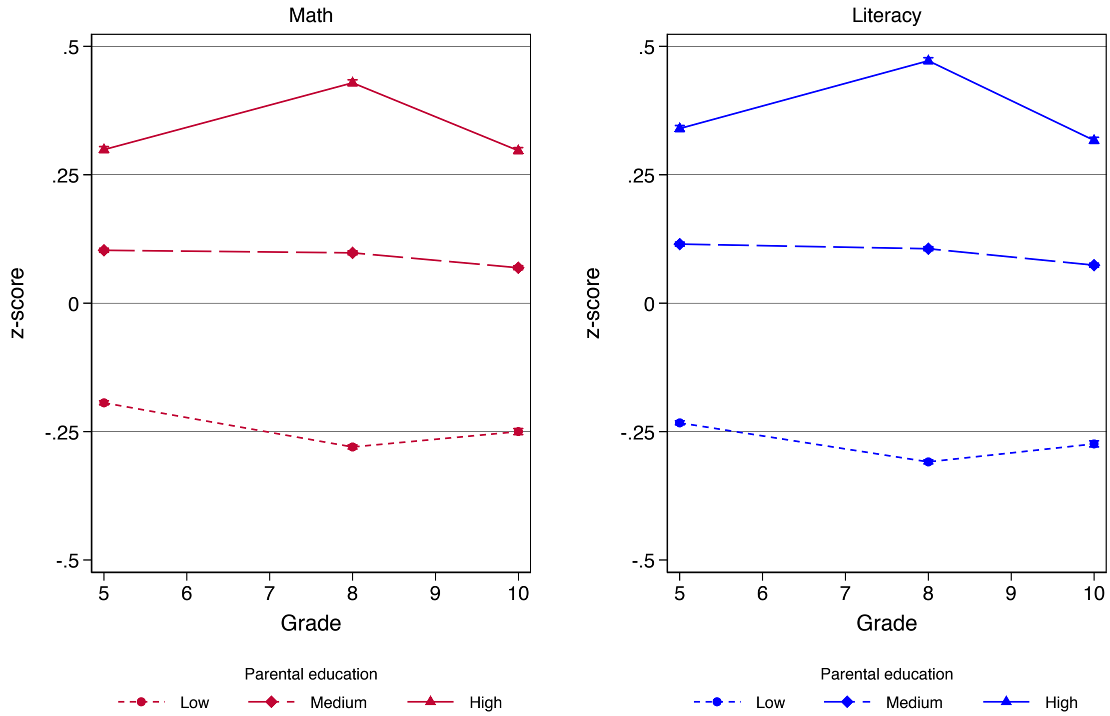


Figure 1 Average educational performance in math and literacy by parental education (adjusted for age).

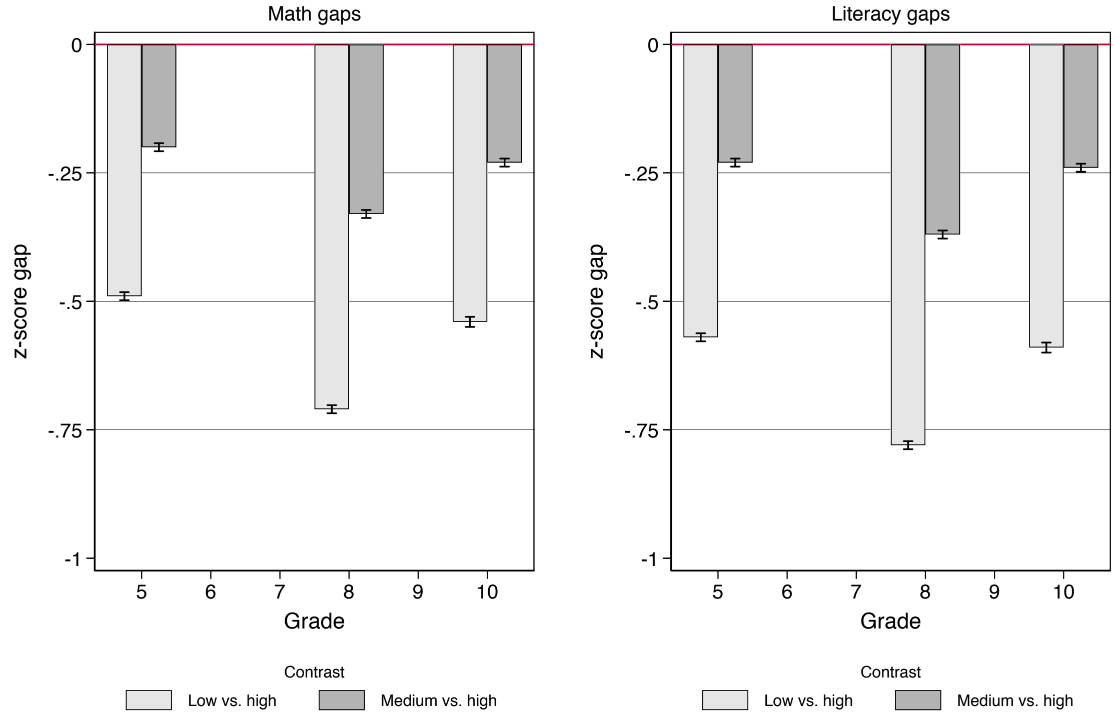


Figure 2 Achievement gaps by parental education (adjusted for age, 'high' as reference).

6.4.2 Achievement gaps by migration status

Next, we assessed achievement gaps by migration status of students. Figure 3 and Figure 4 illustrate results on migrant-native gaps and their trajectory over grades (like for parental education all analyses adjusted for age). Compared to native students, a substantial disadvantage is visible for students with migration background across all grades although gaps slightly reduce for the Grade 10 population. Furthermore, migrant-native gaps are larger for literacy than for math; for example, while the math gap in Grade 8 is about a third of a SD (–.33) to the disadvantage of migrant students, the literacy gap amounts to almost a half of a SD (–.49).

In additional models that modelled a linear trend in z-scores, we found for Math a statistically significant increasing trend over time for immigrants (linear trend coefficient $\beta_t = .052$; $p < .001$) while no significant trend rises for natives. Moreover, despite the fact that Italian data are non-longitudinal in their own nature, we try to assess if and how these three achievement gaps change over time – adopting an empirical strategy *à la difference in difference* – by considering grade/wave as a continuous variable and including in the econometric model the interaction term between wave and migrant status (*wave*×*migrant*), whose coefficient (beta) is the one of primary interest. In this model, we found a statistically significant reduction of the skill gap between the two groups over time (p-value=.0012). As regards literacy we found no statistically significant reduction of the gap between the two groups over time (p-value=.221) and only a non-significant increase of z- scores over grades for migrants.

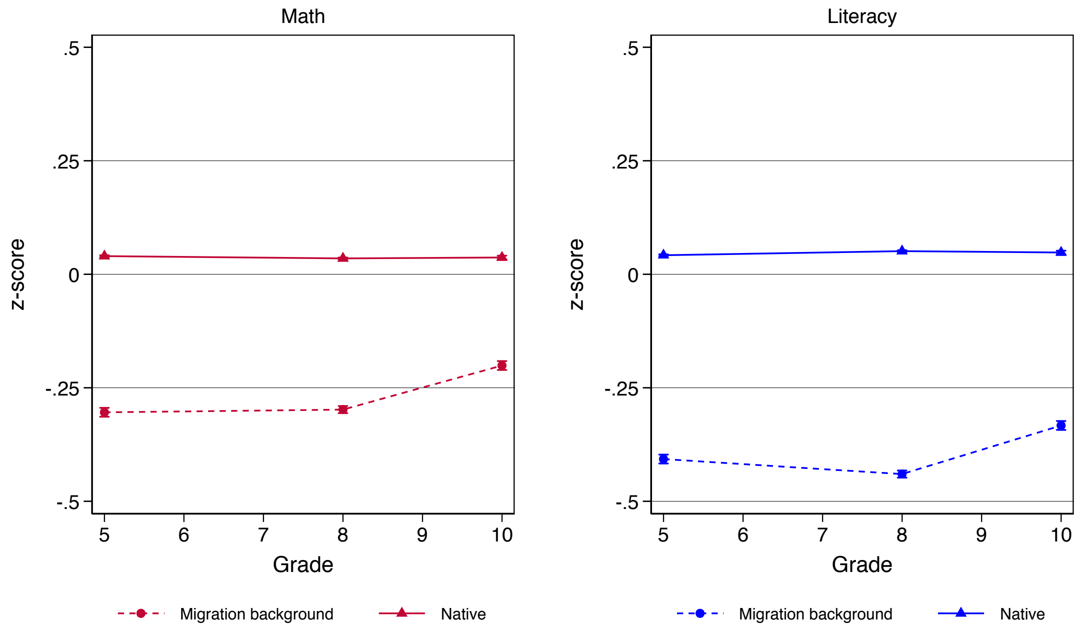


Figure 3 Average performance by migration status (adjusted for age).

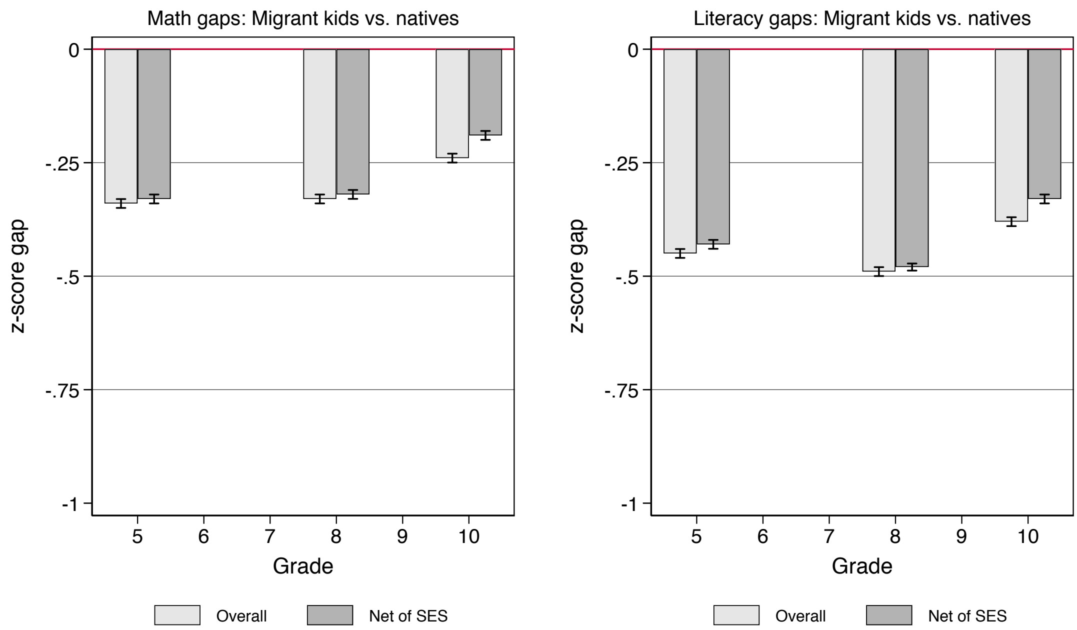


Figure 4 Migrant versus native gaps in achievement overall and after controlling for parental education (adjusted for age, ‘native’ as reference).

Moreover, we how much of the migrant-native gaps remain after accounting for the effects of parental education. Figure 4 shows that gaps only slightly reduce after additionally accounting for parental education (see ‘Net of SES’ bars). We observe a significant reduction of the achievement gap between migrants and natives especially from Grade 5 to Grade 10 (a reduction of .14 SD for math and .10 SD for literacy), even though this is due essentially to the decreasing achievement trajectories for the native students only over the considered grades. To this end, we compare the first-to-last grade differences for the two grade populations: the migrant-native gap from Grade 5 to 10 shows a statistically significant reduction of .14 SD in math ($p < .0001$). At the same time, for literacy, we find a less pronounced but also statistically significant reduction of the achievement gap by .1 SD ($p < .0001$). Finally, we found a statistically significant trend of increasing Math scores (conditional on parental education) for immigrants with respect to natives (coeff. = .02 and $p < .001$). For literacy, however, the trend for migrant children was not statistically significant despite the decreasing migrant-native gap.

6.4.3 First- and second-generation migrants

We further distinguished among migrants between first- and second-generation migrants. As Figure 5 indicates the first-generation migrant versus native gap is significantly larger than the second-generation migrant versus native gap for both domains of math and literacy (all analyses adjusted for parental education and age). Furthermore, we see an apparent reduction of the second-generation migrant versus native gap from Grade 5 to 10. The first-generation migrant versus native gaps increase for Grade 8 but shrink again in Grade 10. In Figure 6, we plotted first- and second-generation migrant versus native gaps that are additionally adjusted for parental education which, however, does not change any conclusion.

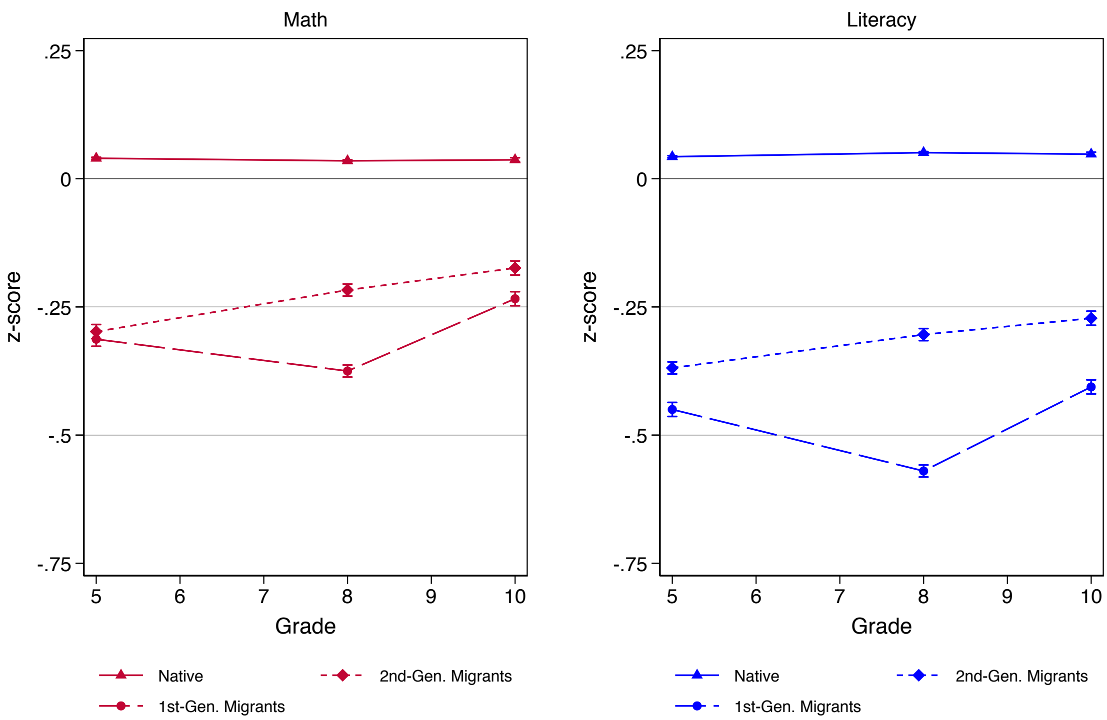


Figure 5 Average performance by migration status differentiated by first- and second-generation migrants (adjusted for age).

6.4.4 Migrant-native by parental education

Do migrant-native gaps vary by levels of parental education? Figure 7 plots average z-scores for migrant and native students by the most extreme groups of parental education (high and low) for both domains under study. This strategy has the advantage to assess more specifically the gaps in each group of interest, instead of using parental education as a control variable in the model as before.

Interestingly, among students from low education backgrounds, we found an impressive reduction of the migrant-native math gap of almost .18 SD ($p < .0001$) between Grade 5 and 10. In stark contrast, the migrant-native gap in math remains stable among children from high education backgrounds. Similarly, for literacy, we found a significant reduction of the migrant-native gap from Grade 5 to 10 (–.12 SD, $p < .0001$) for students from a low educational background, while stable gaps for students from high education backgrounds.

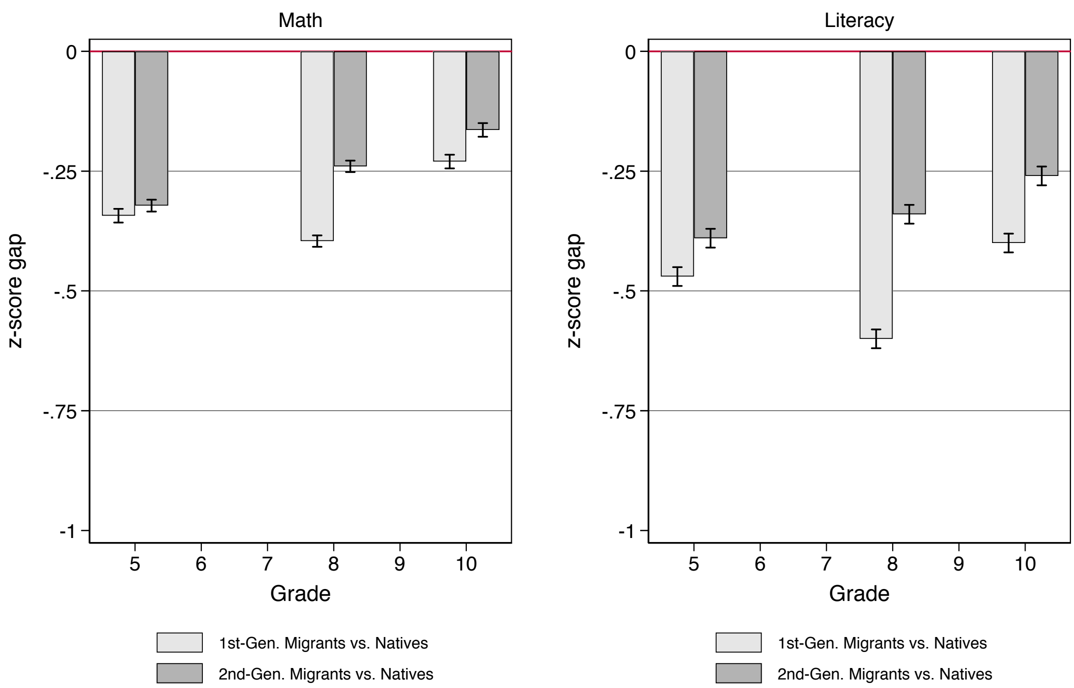


Figure 6 Gaps for first- and second-generation migrants versus native students (adjusted for parental education and age, 'native' as reference).

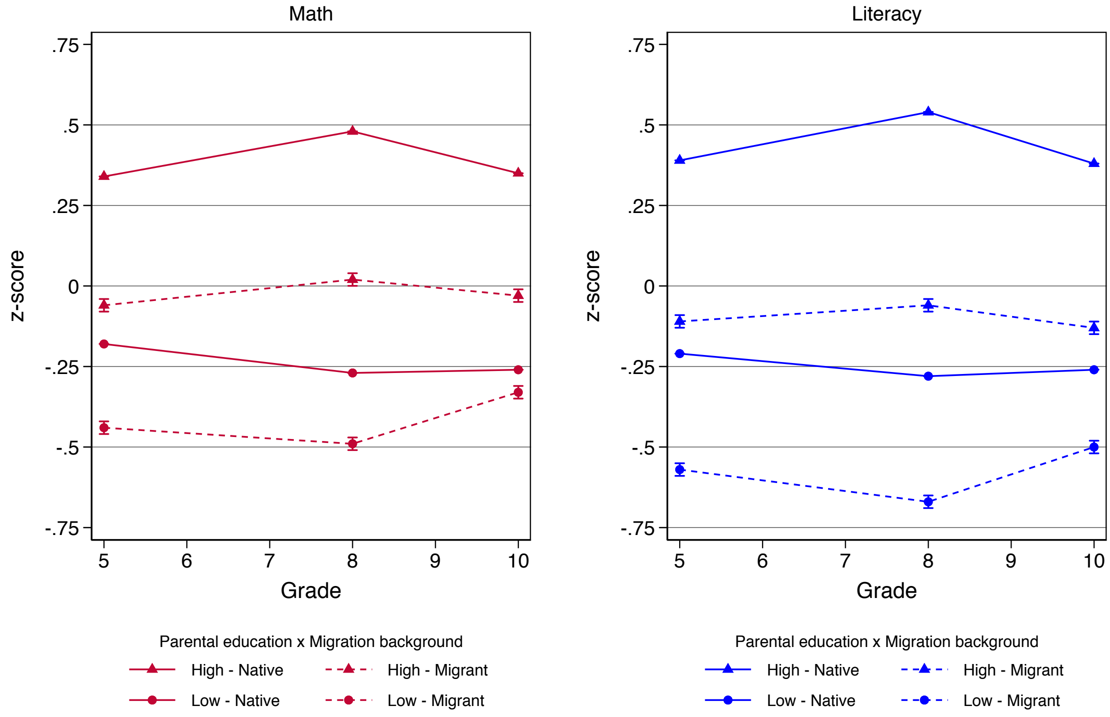


Figure 7 Migrants versus native gaps by parental education (adjusted for age).

6.4.5 Migrant-native gaps by gender

Finally, we inspected the moderating role of gender for migrant-native gaps. Figure 8 shows migrant-native gaps by gender, all adjusted for age and parental education. Overall, we observe a reduction of the migrant-native gap in math over the three grades for both boys and girls (left panel in Figure 7). For boys, the migrant-native gap reduced from Grade 5 to Grade 10 by .12 SD ($p < .0001$). In comparison, for girls, the reduction was more pronounced by .16 SD ($p < .0001$). However, within the populations of native students as well as students with migration background the gender gap in math scores was increasing over grades (to the advantage of boys).

The right panel in Figure 7 shows the same analysis for literacy scores. Although migrant-native gaps were reducing too, the reduction was more pronounced among boys (–.12 SD, $p < .0001$) than for girls (–.07 SD, $p < .0001$). Reversed to math, we see a female advantage for literacy which is growing for both populations of native and migrant students.

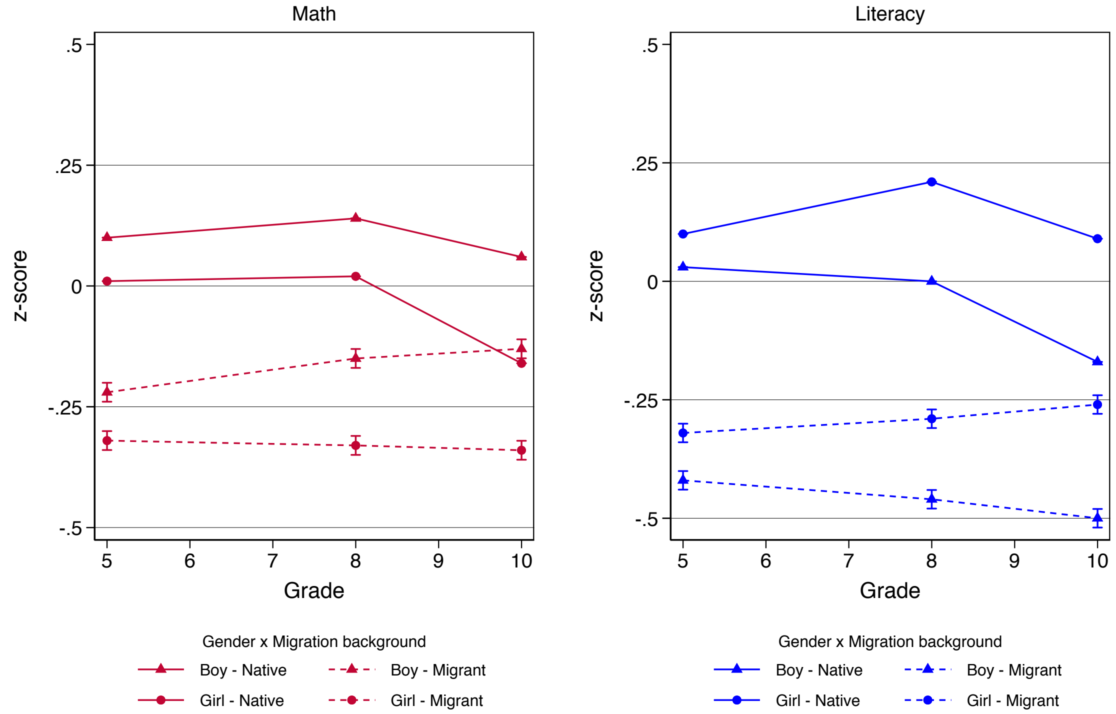


Figure 8 Migrants versus native gaps by gender (adjusted for parental education and age).

6.5 Conclusions

This chapter aimed to study the evolution of social and migration-related achievement gaps in Italian primary and secondary education. Our empirical analyses exploited census data on Italian students' educational achievement in math and literacy, two foundational domains of educational achievement. Using this population data, we constructed a pseudo-panel design that started with a cohort of students who were observed in Grade 5 in 2010–11 (last year of primary schooling) and subsequently followed that cohort up over Grade 8 (third year of lower secondary schooling) and Grade 10 (second year of upper secondary schooling).

Family SES, in our case measured by parental education, plays a substantial role in shaping the educational achievement of students in Italy. In Grade 5, when students were about 10 years old, students from lowly educated parents scored roughly 50% of a SD lower than students from highly educated parents. This gap rose up to roughly 75% of SD in Grade 8 (age about 13) and, albeit a bit smaller, remained substantial in Grade 10 (age about 15).

In contrast to that, we observed migration-related gaps to shrink over time. Moreover, we found that first-generation migrants are most disadvantaged in educational achievement and that educational disadvantage of migrants, in general, is larger for literacy than for math. More refined analyses inspecting the intersection of parental education and migration status revealed that migrant-native gaps in education are smaller for students from lower rather than higher educated parents. Also, we found that convergence of achievement levels among migrant and non-migrant students seem to be driven by the lower SES groups. In contrast, in the higher SES groups, migrant-native gaps remain astoundingly stable.

Finally, gender played a role in achievement. Female students are better in literacy and male students better in math, and the gender gap grows over time for both domains. However, we found an interesting gender interaction with respect to the development of migrants' educational disadvantage. From Grade 5 to Grade 10, male and not female migrant students gained considerable grounds in math while female and not male migrant students gained ground in literacy. Thus, within both genders, migrant-native gaps were shrinking over the observed time window.

7 CONCLUSIONS

Lessons Learned from Five Countries: Summary and Policy Implications

Giampiero Passaretta♠ & Jan Skopek♠
♠Trinity College Dublin

This report of ISOTIS Work Package 1 presented a comprehensive longitudinal study of social and migration gaps in educational achievement in a European-wide comparative perspective. By analysing the evolution of achievement gaps in children from infancy and preschool age up to end of compulsory schooling for Germany, Netherlands, Norway, the United Kingdom, and Italy, our study provides an important contribution to an accumulating body of longitudinal research that is concerned with the roots of educational inequality along socio-economic and migration-related lines. The five countries included represent a theoretically and pragmatically intriguing selection of cases because they exhibit partly stark differences in institutional settings related to early childhood education and care, the organisation of schooling, as well as the overall societal system of stratification.

All country studies drew upon recent high-quality cohort data to address two key research questions: (1) When do social and migration/ethnicity gaps in children's achievement arise and how do they evolve when children are growing up and navigating from infancy to preschool, from preschool to school, and from primary to secondary school? And, (2), to which degree are social and migration-related inequalities in school-age achievement already determined by early inequalities well-established before children enter school? To address these questions, all studies adopted a relative approach that measures inequality in skills at various stages in the early educational career. Detailed analyses at the backdrop of the two guiding research questions are documented in five country chapters authored by researchers involved in ISOTIS Working Package 1.

In the following, we are summarising the main findings related to inequality in children's cognitive and educational achievement. The discussion is split into findings related to inequality by socio-economic status of the origin family and findings related to the immigration or ethnic background of children. The chapter will conclude with several key messages and policy implications.

7.1 Socio-economic status inequality in achievement

All country studies generally reported substantial inequality in educational achievements by the socio-economic status of children's family of origin. Children from high-income families and parents with a high level of education perform consistently better than children from less affluent families and whose parents have less educational resources.

Importantly, these socially-determined gaps are already visible in the very early years of life, tend to increase steadily over infancy, and are well-established even before children enter primary education. For example, at age 5, the average difference in the performance of children from high- and low-educated parents ranges from around .3 SD in Norway (vocabulary) to .6 SD in the UK (composite ability index), about .7 SD in the Netherlands (language)[1], and 1 SD in Germany (composite ability index).[2] Although these cross-country differences are not strictly comparable due to the partly different nature of tests as well as different sampling approaches and data collection methods, our results tentatively suggest that inequalities by socio-economic status are more pronounced in Germany and the Netherlands compared to the liberal welfare regime of the UK, while being substantially lower in the social-democratic regime of Norway.

After increasing over early childhood, SES gaps in achievements remain quite stable and only slightly increase after school entry and throughout primary education. The substantial stability of SES gaps is also apparent over the years of secondary education in almost all countries (with exception of Italy). For example, in Germany, featuring one of the most infamously stratifying systems of secondary schooling, the performance gap between children from low and highly educated parents increases by only 10% from preschool (age 5) to the end of primary schooling (age 10), and remains unchanged until the end of upper secondary education (age 15-16).[3] The pattern of stability over the school years is even more apparent when looking at differences by household income, thus leading us to conclude that socio-economically determined gaps in educational achievements remain stable and, if anything, only slightly increase over schooling in the context of Germany.

In the UK a similar pattern was found: achievement gaps by parental education (high vs low) increase by 8% and 23% moving from age 5 to the end of primary and lower secondary school (age 14), respectively.[4] What is more, in the UK, SES gaps measured by differences between high- and low-income families even slightly decrease as children navigate throughout the education system. The Dutch case revealed a slightly different pattern, however. Differences in the average achievements of children from high- and low-educated parents increase moving from the preschool age (5) to the end of primary education (age 11) but reduce again over lower secondary education: at age 14, the high-low SES gap is 10% higher compared to the preschool gap.[5] Finally, the analysis on Norway showed evidence that points towards a persistence of SES gaps in achievement over the educational career, even if it is more problematic to quantify the percentage changes over the institutional periods due to methodological differences compared to the other country-cases. The only exception to this pattern of stability is Italy, where SES-based inequalities in educational achievement seem to increase at the transition from primary to secondary schooling, while decrease again for students who remain in upper secondary education.

Notwithstanding subtle dissimilarities of findings across countries, we found considerable similarities in the pattern of results despite clear institutional differences in the structure of national education systems and overall welfare arrangements. Moreover, cross-country variations in findings did not covary systematically with the institutional features of the countries. For example, in none of the country-cases SES gaps in achievement were found to increase sharply over

---

[1] Average across measures from Pre-Cool and Cool data.

[2] Results based on models controlling for migration background (for the case of UK, additional analyses).

[3] Results based on the composite ability index.

[4] Results based on the composite index and controlling for migration background (additional analyses, not included in the UK chapter).

[5] Results based on the CITO scores from the COOL data only.

secondary schooling; not even in the context of the strongly stratifying education systems of the Netherlands and especially Germany which both track students at early ages (age 10 and 12, respectively). While there is no sign of a substantial increase in SES-related achievement inequality after students allocate to different school tracks in Germany, SES inequality seems even to decrease after the track allocation in the Netherlands. Overall, our results point towards the importance of the early years of life for the emergence of social stratification in the learning outcomes of children.

The significance of the early years for the social stratification of learning outcomes in school life was confirmed by several chapters that quantified the proportion of SES inequalities in school that are a direct consequence of inequalities established in the preschool period. Findings obtained from Germany, Netherlands and the UK suggest that at least 50% of SES inequalities in school-age achievement are explained by disparities that are well-settled before children enter primary schooling. There is some country-heterogeneity in this proportion, however. When measuring SES by parental education, we estimated that achievement inequalities among children in preschool age (5) explain more than two-thirds (70%) of inequalities at age 9 in Germany, and around half of the disparities in the end of primary education (age 11) in the Netherlands (50%) and the UK (55%). Inequalities determined by families' financial capacity in the early years account for around 70% and 55–60% of primary school inequalities in the UK and Germany, respectively. Whereas is not possible to offer an exact estimation due to methodological differences in the analyses of the Norwegian context, the Norwegian results suggest that a significant part of later SES gaps in achievement is explained by inequalities that accrued in the very early years of life.

Taken together, our analyses suggest that a remarkable part of SES inequalities in achievement accumulated over infancy and childhood is carried over into the school system. Therefore, preschool inequalities seem to be responsible for a large part of the disparities observed in classrooms. These results have important implications from a policy perspective. They suggest that the most effective way of reducing SES disparities in school achievements is promoting policies that favour a reduction of skill gaps before the school entry. Whether directed towards the expansion of ECEC systems and preschool education facilities or the direct support of parents and children in their family environments, these policies – if effective – have the potential to reduce plenty of social disparities in achievement in the classrooms.

Aside the importance of the early years, factors related to SES of families continue to shape children's achievement in school, at least to some extent. Even when entering primary education with the same level of achievement, children from socio-economically worse-off families fall behind their peers from better-off families over the school career. However, our comparative study revealed some interesting cross-country variation in the substantive 'function' of this additional SES effect. In Germany, the additional role of SES concentrates among poor preschool achievers, thus helping high-SES children with poor preschool performances to partly recover over their school careers. Conversely, high- and low-SES children starting with top preschool performances seem to perform equally over schooling. Contrasting cases are the Netherlands and the UK. Here, the SES effect was found to be concentrated among top preschool achievers: when starting school with the same high intellectual capacity, low-SES children are less able to hold the top positions compared to high-SES kids. However, although to a lower extent, the additional advantage of high-SES kids is visible even at low-end of the distribution of preschool achievements in the latter two countries. Therefore, family SES seems to ensure compensation of poor initial performances in the German context, while both compensating for poor starts and especially boosting the educational careers of initially well-performing kids in the Netherlands and the UK.

7.2 Ethnic and migration-related inequality in achievement

Compared to the fairly consistent picture for SES-related achievement inequality, our study revealed more cross-country heterogeneity in findings on ethnic and migration-related inequalities in educational achievements. In most of the countries, migration-related disparities in achievement are sharp in the early years of life but decrease as children navigate through the education systems substantially. Generally speaking, differences in the performances of migrant and non-migrants are more pronounced in verbal-related competencies compared to other competence domains, such as math or quantitative skills. There are significant country differences, however.

In the UK and the Netherlands, migrant penalties close completely over schooling, although migrant children catch up to children of natives faster in the UK (already at age 7 after 2 years of schooling) than in the Netherlands (at the end primary or over secondary schooling at the latest). Socio-economic differences between migrant and non-migrant families do, however, explain only little of the early disadvantages children of immigrants are facing in the UK. Language culture at home was identified as the primary factory that explained migrant children's disadvantage in the UK. Italy is a different case: while showing a tendency of reducing migrant-gaps at the transition from primary to lower secondary school, the migrant-native gap remains substantial by the age of 15. Moreover, in the Italian context, only a minority of the literacy penalty of first (33%) and second-generation migrants (5%) is explained by their generally lower SES compared to natives.[6] In Germany, we saw that the evolution of migration-related achievement inequalities was more dependent on the cognitive domain considered. While in some domains, such as math or reading, inequalities remained rather stable over time, the average difference in the mastery of German vocabulary between migrants and natives seemed to reduce as children progress through their school career, although migrants still lag behind at the end of upper secondary education. Compared to other countries, such as the UK and Italy, a striking difference of the German context was that the lower SES of migrants' families explained a large proportion (around 50%) of migrants' disadvantages for nearly all domains. Finally, findings obtained for Norway were more contrasting and more dependent on the dataset used for the analyses. However, the evidence from the larger dataset (MoBa) available suggests a substantial and persistent penalty for migrant children in the mastery of language over early childhood and the early years of schooling.

Our comparative study could also shed some light on the dynamics of inequalities among some of the specific groups of migrants targeted by ISOTIS, such children with Turkish and Moroccan background. However, such analyses were limited to Germany and Netherlands due to the unavailability of the information about these specific ethnic groups in most datasets. Our results consistently point towards a strong disadvantage of children with a Turkish background. Turkish perform substantially lower than natives and even other migrant groups in both countries.

[6] The proportions reported are the averages over Grade 5 and Grade 8.

Despite this commonality, there are cross-country differences in the evolution of such disparities along the educational cycle. While in the Netherlands the considerable Turkish-native gap reduces substantially in all domains and even disappears (in the case of math) by the end of lower secondary education, there is no sign of catching-up over schooling for Turkish children in Germany. Instead, the relative position of Turkish compared to natives, and even to other groups of migrants, seems even to deteriorate slightly as children navigate through the German school system. It is also interesting to note that, in the only country that allows such analyses (the Netherlands), the patterns of results for children with a Moroccan background is similar to that found for Turkish. While being somewhat smaller, the disadvantage of Moroccan tends to decrease considerably over schooling and vanishes entirely in an important competence domain such as mathematics.

Unfortunately, due to data limitations, it was not possible to identify ISOTIS target groups in other countries. Still, the UK chapter provided some compelling evidence concerning the dynamics of inequalities for prominent minority groups in terms of their ethnic and religious background. Three findings are of particular relevance. First, the achievements of black minorities – such as black African, black Caribbean, black British, and Pakistani or Bangladeshi – are considerably lower compared to white children. As regards religious minorities, Muslims seem to be the most disadvantaged group in terms of achievement. Second, the UK chapter found significant interactions between ethnicity and migration status: non-white children with a migration background face the strongest disadvantages at the beginning of their education journey. Third, it is striking that the achievement penalties of all ethnic minorities disappear upon the school entry (except for the group of Pakistani/Bangladeshi), while the penalty of Muslims substantially reduces. All in all, while ethnic minority disadvantages are substantial in the early years, a large part of those disadvantages vanish at a fast pace once children enter the education system.

Although differences in the average performance of migrants and natives tend to reduce and even disappear over schooling in most countries, inequalities in the early years are still crucial in determining migration disparities observed among children in school age. In Germany and the Netherlands, the entire migrants' penalty in primary school achievement is attributable to migrants' lower performances in preschool age. What is more, in these two conservative welfare regimes featuring a long-standing tradition of 'guest workers', children with a migration background even outperform natives when having the same level of achievement in the preschool age. Therefore, it can be concluded that the over-proportional gain of migrants beyond preschool age is a critical compensatory mechanism acting towards decreasing inequality or at least counterbalancing the growth of the migrant-native gap in achievement over the school career. Without these over-proportional gains, migrants' penalties would indeed decrease at lower rates (as observed) in the Netherlands and even be increasing rather than remaining stable in Germany.

The most significant over-proportional gains of migrants and other minority groups over the school career were found in the UK. Here, the over-proportional gains of minorities are responsible for almost complete eradication of migration-related gaps in achievement over the educational careers. The astounding catching-up process of minority groups in the UK implies that, when minority children had started at the same level at school entry age, they would have out-performed majority children in school. Hence, the reason why on average minority children are not outperforming majority children in school must be sought in their relative disadvantage in the early years.

Taken together, we found strong evidence for over-proportional achievement gains for children of immigrants (and other minority groups) at the transition from the preschool to the school period, which avoid increasing migration-related inequalities in Germany, lead to a sharp reduction of inequality in Netherlands, and an almost complete catching-up over the educational career in the UK. Finally, the over-proportional gains of migrants acting toward compensation of initial migrant-native inequalities concentrate among top preschool achievers in Germany and poor preschool achievers in both the UK and the Netherlands.

7.3 Key messages and policy implications

The early years of life (before children enter school) are formative for patterns of inequality observed in school age, and this holds for achievement inequality both by socio-economic and migration status. Socio-economic and migration-related achievement gaps in school are therefore rooted substantially in the early years.

7.3.1 Socio-economic status

Key messages

- The major part of SES gaps in educational achievement in school is attributable to children from high and low SES starting unequally into school life.
- A minor part of SES gaps in school is a direct consequence of low-SES children falling behind their high-SES peers starting equally into school life.

Policy implications

- Preschool-age interventions have a higher potential for reducing achievement inequality in school compared to school-age interventions. Hence, supporting children from socio-economically disadvantaged family backgrounds in the early years is the most effective way to reduce achievement gaps in school.
- Although potentially less effective, school-age interventions can help to reduce SES inequalities in the classrooms. The effectiveness of school-age interventions can differ between countries (examples below).
    a) In Germany, school-age interventions are more effective when targeted towards low-SES children with poor performance at school start.
    b) In the UK and the Netherlands, school-age interventions are more effective when targeted towards low-SES children with high achievement level at school start.

7.3.2 Migration background

Key messages

- Children with a migration background enter school with a substantial disadvantage but enjoy over-proportional achievement gains in school.
- The over-proportional gains of migrants in school compensate for the poor performances of migrants in the preschool age.

Policy implications

- Reducing migration-related inequality in preschool age have the potential to eradicate migrants' penalties in school age entirely (and even contribute to emerging migrants' premiums in school).

Finally, we would like to clarify that based on our study we cannot evaluate which kind of policy interventions would be most useful to reduce social and migration-related achievement gaps among children. However, as we tried to highlight based on our findings, policy interventions could be most beneficial if they target to improve the situation of disadvantaged children in the early years before school. Nonetheless, we would like to remind that possible preschool- and school-age interventions are by no means restricted to policies linking to the ECEC or school systems – such as policies expanding ECEC education or policies improving the quality of ECEC services or schools. Promising interventions could also target the home or neighbourhood situation of disadvantaged families through appropriate financial and social support.

8 APPENDIX

Several chapters in this report refer to additional and supplementary analyses which are provided in a separate appendix that is available online at http://www.isotis.org.

9 REFERENCES

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2017). *When should you adjust standard errors for clustering?* NBER Working Paper No. w24003.

Akgündüz, Y. E., & Heijnen, S. (2018). Impact of funding targeted pre-school interventions on school readiness: Evidence from the Netherlands. *De Economist*, *166*(2), 155-178.

Akgündüz, Y. E., & Plantenga, J. (2014a). Childcare in the Netherlands: Lessons in privatisation. *European Early Childhood Education Research Journal*, *22*(3), 379-385.

Akgündüz, Y. E., & Plantenga, J. (2014b). Equal access to high-quality childcare in the Netherlands. In L. Gambaro, K. Stewart, & J. Waldfogel (Eds.), *An Equal Start? Providing Quality Early Education and Care for Disadvantaged Children:* 101–120. Bristol, UK: Policy Press.

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2004). Schools, achievement, and inequality: A seasonal perspective. *Summer Learning: Research, Policies, and Programs*, *23*(2), 25–51.

Allmendinger, J. (1989). Educational systems and labor market outcomes. *European Sociological Review*, *5*(3), 231–250.

Arts, W. A., & Gelissen, J. (2002). Three worlds of welfare capitalism or more? A state-of-the-art report. *Journal of European Social Policy*, *12*(2), 137–158.

Barone, C. (2006). Cultural capital, ambition and the explanation of inequalities in learning outcomes: A comparative analysis. *Sociology*, *40*(6), 1039–1058.

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*(4), 1347–1367.

Becker, B., & Schober, P. S. (2017). Not Just Any Child Care Center? Social and Ethnic Disparities in the Use of Early Education Institutions With a Beneficial Learning Environment. *Early Education and Development*, *28*(8), 1011–1034.

Benavot, A., & Resnik, J. (2004). Lessons from the Past: A Comparative Socio-Historical Analysis of Primary and Secondary Education. In A. Benavot, J. Resnik, & J. Corrales (Eds.), *Global Educational Expansion. Historical Legacies and Political Obstacles*. Cambridge: American Academic of Arts and Sciences.

Bettendorf, L. J., Jongen, E. L., & Muller, P. (2015). Childcare subsidies and labour supply—Evidence from a large Dutch reform. *Labour Economics*, *36*, 112-123.

Bishop, D. V. M. (2003). *The children's communication checklist: CCC-2*. London: Harcourt Assessment.

Blossfeld, H. P., Roßbach, H. G., & von Maurice, J. (2011). Editorial. *Zeitschrift Für Erziehungswissenschaft*, *14*(S2), 1–4.

Blossfeld, H. P., Buchholz, S., Skopek, J., & Triventi, M. (Eds.). (2016). *Models of Secondary Education and Social Inequality: An International Comparison*. Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

BMFSFJ. (2015). *Fünfter Bericht zur Evaluation des Kinderförderungsgesetzes*.

Bodovsky, K., & Youn, M. (2012). Students' Mathematics Learning from Kindergarten through 8th Grade: The Long-Term Influence of School Readiness. *International Journal of Sociology of Education*, *1*(2), 97–122.

Bol, T., & van de Werfhorst, H. G. (2011). Measuring Educational Institutional Diversity: External

Differentiation , Vocational Orientation and Standardization Thijs Bol. *Amsterdam Centre for Inequality Studies Working Paper Series*, *1*(November).

Boterman, W. R. (2019). School Segregation in the Free School Choice Context of Dutch Cities. In X. Bonal, & C. Bellei (Eds.), *Understanding School Segregation: Patterns, Causes and Consequences of Spatial Inequalities in Education*: 155-178. London: Bloomsbury Academic.

Bratti M., Checchi D. & Filippin, A. (2007). Territorial Differences in Italian Students' Mathematical Competencies: Evidence from PISA 2003. *IZA Discussion Papers*, 2603).

Brunello, G., & Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic policy*, *22*(52), 782-861.

Bradbury, B., Corak, M., Waldfogel, J., & Washbrook, E. (2015). *Too many children left behind. The U.S. achievement gap in comparative perspective*. New York: Russel Sage Foundation.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic Status and Child Development. *Annual Review of Psychology*, *53*(1), 371–399.

Broekhuizen, M., Ereky-Stevens, K., Wolf, K., Francot, R., Moser, T., Anders, Y., & Leseman, P. (2018). *Technical report parent structured interview study. Instrument development, descriptive results, and showcases.*

Buchholz, S., Skopek, J., Zielonka, M., Ditton, H., Wohlkinger, F., & Schier, A. (2016). Secondary school differentiation and inequality of educational opportunity in Germany. In H. Blossfeld, S. Buchholz, J. Skopek, & M. Triventi (Eds.), *Models of Secondary Education and Social Inequality. An International Comparison* (pp. 79–92). Celtenham, UK: Edward Elgar.

Bukodi, E., & Goldthorpe, J. H. (2011). Class origins, education and occupational attainment in Britain: Secular trends or cohort-specific effects? *European Societies*, *13*(3), 347–375.

Caro, D. H., Mcdonald, J. T., Willms, J. D., Caro, D. H., Mcdonald, J. T., & Willms, J. D. (2016). Socio - economic Status and Academic Achievement Trajectories from Childhood to Adolescence. *Canadian Journal of Education*, *32*(3), 558–590.

Checchi, D., Ichino, A., & Rustichini, A. (1999). More equal but less mobile? Education financing and intergenerational mobility in Italy and in the US. *Journal of public economics*, 74(3), 351-393.

Checchi, D. (2003). *The Italian educational system: family background and social stratification*. Monitoring Italy, ISAE Conference, Rome.

Checchi, D., & Flabbi, L. (2006). Intergenerational mobility and schooling decisions in Italy and Germany. *IZA Discussion Paper*, *2876*.

Checchi, D. (2010). Percorsi scolastici e origini sociali nella scuola italiana. *Politica Economica*, 26(3), 359-388.

Checchi, D. (2010). Educational achievements and social origins in Italy. *Politica economica*, (3), 359-388.

Checchi, D., & Verzillo, S. (2017). The Role of PISA in Regional and Private/Public Debates in Italy. Volante L.(a cira di), *The PISA Effect on Global Educational Governance*, 127-148.

Chmielewski, A. K., & Reardon, S. F. (2016). Patterns of Cross-National Variation in the Association Between Income and Academic Achievement. *AERA Open, 2*(3).

Conger, R. D., Conger, K. J., & Martin, M. J. (2010). Socioeconomic status, family processes,and individual development. *Journal of Marriage and Family*, *72*(3), 685–704.

Conger, R. D., & Donnellan, M. B. (2007). An Interactionist Perspective on the Socioeconomic Context of Human Development. *Annual Review of Psychology*, *58*(1), 175–199.

Contini, D., & Scagni A. (2013). Social-Origin Inequalities in Educational Careers in Italy. In: M. Jackson (Ed.), *Determined to Succeed? Performance versus Choice in Educational Attainment* (pp. 149-227). Standford: Stanford University Press.

Contini, D., & Triventi, M. (2016). Between formal openness and stratification in secondary education: Implications for social inequalities in Italy. In H.-P. Blossfeld, S. Buchholz, J. Skopek, & M. Triventi (Eds.), *Models of Secondary Education and Social Inequality: An International Comparison* (p.305-322). Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Crozier, G., & Davies, J. (2007). Hard to reach parents or hard to reach schools? A discussion of home-school relations, with particular reference to Bangladeshi and Pakistani parents. *British Educational Research Journal*, *33*(3), 295–313.

Dämmrich, J., & Triventi, M. (2018). The dynamics of social inequalities in cognitive-related competencies along the early life course–A comparative study. *International Journal of Educational Research*, *88*(1), 73–84.

Dearing, E., Zachrisson, H. D., Mykletun, A., & Toppelberg, C. O. (2018). Estimating the consequences of Norway's national scale-up of early childhood education and care (beginning in infancy) for early language skills. *AERA Open, 4*(1), 1-16.

Deutsches Jugendinstitut Dortmunder Arbeitsstelle Kinder- und Jugendhilfestatistik, X. (2008). *Zahlenspiegel 2007. Kindertagesbetreuung im Spiegel der Statistik.*

Deutsches Pisa-Konsortium. (2001). *Pisa 2000*.

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*(1), 271–297.

Domina, T. (2005). Leveling the Home Advantage : Assessing the Effectiveness of Parental Involvement in Elementary School. *Sociology of Education*, *78*, 233–249.

Doyle, O., Harmon, C. P., Heckman, J. J., & Tremblay, R. E. (2009). Investing in early human development: timing and economic efficiency. *Economics & Human Biology*, *7*(1), 1-6.

Downey, D. B., & Condron, D. J. (2016). Fifty Years since the Coleman Report: Rethinking the Relationship between Schools and Inequality. *Sociology of Education*, *89*(3), 207–220.

Downey, D. B., von Hippel, P. T., & Broh, B. A. (2004). Are Schools the Great Equalizer ? Cognitive Inequality during the Summer Months and the School Year. *American Sociological Review*, *69*(5), 613–635.

Driessen, G., Mulder, L., Ledoux, G., Roeleveld, J., & van der Veen, I. (2009). Cohortonderzoek COOL$^{5–18}$. Technisch rapport basisonderwijs, eerste meting 2007/08. Nijmegen: ITS / Amsterdam: SCO-Kohnstamm Instituut.

Duncan, G. J., & Magnuson, K. A. (2003). Off with Hollingshead: Socioeconomic resources, parenting, and child development. In M. H. Bornstein & R. H. Bradley (Eds.), *Socioeconomic status, parenting, and child development* (pp. 83–106). Mahwah, New Jersey: Lawrence Erlbaum.

Duncan, G. J., Magnuson, K., & Votruba-Drzal, E. (2015). Children and socioeconomic status. In *Handbook of Child Psychology and Developmental Science: Ecological Settings and Processes* (pp. 534–573).

Education Council (2018). *Stand van educatief Nederland 2018*. Onderwijsraad: The Hague.

Elango, S., García, J. L., Heckman, J. J., & Hojman, A. (2015). *Early childhood education*. NBER Working Paper No. 21766.

Ermisch, J. (2008). Origins of Social Immobility and Inequality: Parenting and Early Child Development, (205), 62–71.

Ertl, H. (2006). Educational standards and the changing discourse on education: the reception and consequences of the PISA study in Germany. *Oxford Review of Education ISSN:*, *32*(5), 619–634.

Esping-Andersen, G. (1990). *The Three Worlds of Welfare Capitalism*. Oxford: Polity Press.

Eurofound. (2017). *Income inequalities and employment patterns in Europe before and after the Great Recession*.

European-Commission/EACEA/Eurydice/Eurostat. (2014). *Key Data on Early Childhood Education and Care in Europe*. Luxembourg. https://doi.org/10.1007/978-1-4613-3479-8.

Fan, X., & Chen, M. (2001). Parental Involvement and Studnet's Academic Achievement: A Meta-Analysis. *Educational Psychology Review*, *13*(1), 1–22.

Feinstein, L. (2003). Inequality in the early cognitive development of British children in the 1970 cohort. *Economica*, *70*(277), 73–97.

Fitzsimons, E. (2017). *Millennium Cohort Study. Sixth Survey 2015-2016. User Guide (First Edition).* London.

Hansen, K. (2014). *Millennium Cohort Study. A Guide to the Datasets (Eighth Edition).* London.

Heeringa, S. G., West, B. T., & Berglund, P. A. (2010). *Applied Survey Data Analysis*. Boca Raton: Taylor & Francis.

Hoffmann, N. I. (2018). Cognitive achievement ofchildren of immigrants: Evidence from the Millennium Cohort Study and the 1970 British Cohort Study. *British Educational Research Journal*, (advanced access).

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, 16(2), 234-248.

Fryer, B. R. G., & Levitt, S. D. (2013). Testing for Racial Differences in the Mental Ability of Young Children. *The American Economic Review*, 103(2), 981-1005.

Gresch, C., Baumert, J., & Maaz, K. (2010). Empfehlungsstatus, Übergangsempfehlung und der Wechsel in die Sekundarstufe I: Bildungsentscheidungen und soziale Ungleichheit. In *Bildungsentscheidungen* (pp. 230–256).

Gresham, F., & Elliott, S. N. (2008). Social skills improvement system (SSIS) rating scales. *Bloomington, MN: Pearson Assessments*.

Gruber, K. H. (2006). The German 'PISA-Shock': some aspects of the extraordinary impact of the OECD's PISA study on the German education system. In *Cross-national Attraction in Education: Accounts from England and Germany*. Symposium Books.

Heckman, J. J., & Mosso, S. (2014). The economics of human development and social mobility. *Annual Review of Economics*, *6*(1), 689-733.

Helbig, M., & Nikolai, R. (2015). *Die Unvergleichbaren. Der Wandel der Schulsysteme in den deutschen Bundesländern seit 1949*. Bad Heilbrunn: Verlag Julius Klinkhardt.

Helland, W. A., & Heimann, M. (2007). Assessment of pragmatic language impairment in children referred to psychiatric services: a pilot study of the Children's Communication Checklist in a Norwegian sample. *Logoped Phoniatr Vocol, 32*(1), 23-30.

Helland, W. A., & Møllerhaug, L. (2002). *CCC-2 Norwegian Adaptation*.

Hoffman, N. (2018). Cognitive achievement of children of immigrants: Evidence from the Millennium Cohort Study and the 1970 British Cohort Study. *British Educational Research Journal*.

Iannelli, C. (2013). The role of the school curriculum in social mobility. *British Journal of Sociology of Education*, *34*(5–6), 907–928.

Inspectorate of Education (2018). *De staat van het onderwijs. Onderwijsverslag 2016/2017*. Inspectie van het Onderwijs: Utrecht.

Inspectorate of Education (2015). *De staat van het onderwijs. Onderwijsverslag 2013/2014*. Inspectie van het Onderwijs: Utrecht.

Janssen, J., Scheltens, F., and Kraemer, J.-M. (2005). *Leerling- en Onderwijsvolgsysteem Rekenen-wiskunde*. Arnhem: Cito.

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. Series A: Statistics in Society*, *176*(4), 887–906.

Jin, W., Muriel, A., & Sibieta, L. (2010). *Subject and course choices at ages 14 and 16 amongst young people in England: Insights from behavioural economics*. *Institute for Fiscal Studies*. London.

Kautz, T., Heckman, J. J., Diris, R., Ter Weel, B., & Borghans, L. (2014). *Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success*. NBER Working Paper No. 20749.

Klein, W. (2005). Alle Länder auf gutem Weg? Erste Ergebnisse des zweiten PISA-Ländervergle-ichs. *Schulmanagement*, *5*, 18–21.

Klein, M., Iannelli, C., & Smyth, E. (2016). School subject choices and social class differences in entry to higher education - Comparing Scotland and Ireland. In H.-P. Blossfeld, S. Buchholz, J. Skopek, & M. Triventi (Eds.), *Models of Secondary Education and Social Inequality. An International Comparison.* (pp. 233–248). Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Kristen, C., & Dollmann, J. (2010). Sekundäre Effekte der ethnischen Herkunft: Kinder aus türkischen Familien am ersten Bildungsübergang. In B. Becker & D. Reimer (Eds.), *Vom Kindergarten bis zur Hochschule* (pp. 117–144). VS Verlag für Sozialwissenschaften.

Koerhuis, I., and Keuning, J. (2011). *Wetenschappelijke Verantwoording van de Toetsen Rekenen voor Kleuters*. Arnhem: Cito.

Ladd, H. F., & Fiske, E. B. (2011). Weighted student funding in the Netherlands: A model for the US?. *Journal of Policy Analysis and Management*, *30*(3), 470-498.

Ladd, H.F., Fiske, E.B., & Ruijs, N.M. (2011). Does Parental Choice Foster Segregated Schools? Insights from the Netherlands. In Berends, M., Cannata, M., & Goldring, E.B. (Eds). *School Choice and School* Improvement:233-253. Cambridge: Harvard Education Press.

Lansink, N., and Hemker, B. (2012). *Wetenschappelijke Verantwoording van de Toetsen Taal voor Kleuters voor Groep 1 en 2 uit het Cito Volgsysteem Primair Onderwijs*. Arnhem: Cito.

Lee, V. E., & Burkam, D. T. (2002). *Inequality at the starting gate: Social background differences in achievement as children begin school*. Washington, D.C.: Economic Policy Institute.

Leseman, P., Mulder, H., Verhagen, J., Broekhuizen, M., van Schaik, S., & Slot, P. (2017). 9. Effectiveness of Dutch targeted preschool education policy for disadvantaged children: Evidence from the pre-COOL study. In H. P. Blossfeld, N. Kulic, J. Skopek & M. Triventi (Eds.), *Childcare, early education, and social inequality – An international perspective* (pp. 173-193). Cheltenham, UK: Edward Elgar.

Leu, H. R., & Schelle, R. (2009). Between education and care? Critical reflections on early childhood policies in Germany. *Early Years*, *29*(1), 5–18.

Leuven, E., Lindahl, M., Oosterbeek, H., & Webbink, D. (2010). Expanding schooling opportunities for 4-year-olds. *Economics of Education Review*, *29*(3), 319-328.

Levels, M., Dronkers, J., & Kraaykamp, G. (2008). Immigrant Children ' s Educational Achievement in Western Countries : Origin , Destination , and Community Effects on Mathematical Performance. *American Sociological Review*, *73*(5), 835–853.

Magnus, P., Birke, C., Vejrup, K., Haugan, A., Alsaker, E., Daltveit, A. K., Knudsen, G. P. (2016). Cohort Profile Update: The Norwegian Mother and Child Cohort Study (MoBa). *Int J Epidemiol*, dyw029.

Magnus, P., Irgens, L. M., Haug, K., Nystad, W., Skjaerven, R., Stoltenberg, C., & MoBa Study, G. (2006). Cohort profile: the Norwegian Mother and Child Cohort Study (MoBa). *Int J Epidemiol, 35*(5).

Magnuson, K. A., Meyers, M. K., Ruhm, C. J., & Waldfogel, J. (2004). Inequality in preschool education and school readiness. *American educational research journal*, *41*(1), 115-157.

Manning, A., & Pischke, J.-S. (2006). *Comprehensive versus Selective Schooling in England and Wales: What Do We Know?* (Discussion Paper No. 2072). *Institute for the Study of Labor*. Bonn.

Manzoni, A., Härkönen, J., & Mayer, K. U. (2014). Moving on? A growth-curve analysis of occupational attainment and career progression patterns in West Germany. *Social Forces*, *92*(4), 1285–1312.

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*(2), 105–128.

McLoyd, V. C. (1998). Socioeconomic Disadvantage and Child Development. *American Psychologist*, *53*(2), 185–204.

McMullin, P., & Kulic, N. (2016). Onwards or upwards? – The role of subject choice and schools in the reproduction of educational ienquality in England. In H.-P. Blossfeld, S. Buchholz, J. Skopek, & M. Triventi (Eds.), *Models of Secondary Education and Social Inequality. An International Comparison.* (pp. 213–231). Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Ministry of Education, Culture and Science (2016). *Key Figures Education*. The Hague, the Netherlands: Ministry of Education, Culture and Science.

Muffels, R., & Luijkx, R. (2008). Labour market mobility and employment security of male employees in Europe: "Trade-off" or "flexicurity"? *Work, Employment and Society*, *22*(2), 221–242.

Mulder, H., Hoofs, H., Verhagen, J., van der Veen, I., & Leseman, P. P. (2014). Psychometric properties and convergent and predictive validity of an executive function test battery for two-year-olds. *Frontiers in Psychology*, *5*, 733.

Nærde, A., Janson, H. T., & Ogden, T. (2014). *BONDS (The Behavior Outlook Norwegian Developmental Study): A Prospective Longitudinal Study of Early Development of Social Competence and Behavior Problems. Oslo, Norway: The Norwegian Center for Child Behavioral Development*.

Neugebauer, M. (2010). Bildungsungleichheit und Grundschulempfehlung beim Übergang auf das Gymnasium: Eine Dekomposition primärer und sekundärer Herkunftseffekte. *Zeitschrift Für Soziologie*, *39*(3), 202–214.

Noailly, J., & Visser, S. (2009). The impact of market forces on child care provision: Insights from the 2005 child care act in the Netherlands. *Journal of Social Policy*, *38*(3), 477-498.

Norway, S. (2014, 2014). Gjennomstrømning i videregående opplæring, 2007-2012 [Completion of secondary education, 2007-2012].

Norway, S. (2015). *Inntekts- og fomuestatistikk for husholdninger [Income an wealth statistics for households]*.

Norway, S. (2018). *ECEC statistics*. Retrieved from Oslo, Norway: Statistics Norway

Oakes, J. M., & Rossi, P. H. (2003). The measurement of SES in health research: current practice and steps toward a new approach, *56*, 769–784.

Oberheumer, P., & Ulich, M. (1997). *Working with Young Children in Europe: Provision and Staff Training*. London: Paul Chapman Publishing.

OECD. (2015). *Starting Strong IV*. Paris: OECD Publishing.

OECD (2016a). Enrolment in childcare and pre-school. OECD Family Database.

OECD (2016b). *Netherlands 2016: Foundations for the Future*. Reviews of National Policies for Education. Paris: OECD Publishing.

OECD. (2016c). *Society at a Glance 2016. OECD Social Indicators.* Paris: OECD Publishing.

OECD. (2017). *Starting Strong 2017: Key OECD Indicators on Early Childhood Education and Care*. Paris: OECD Publishing

OECD. (2018a). *Settling In 2018: Indicators of Immigrant Integration*. Paris: OECD Publishing.

OECD (2018b). Education spending (indicator). OECD Data.

OECD (2018c). *Education at a Glance*. Paris: OECD Publishing.

OECD (2018d). *Adult Education Level (indicator)*. Paris: OECD Publishing.

O'Higgins, N., D'Amato, M., Caroleo, F. E., & Barone, A. (2007). Gone for Good? Determinants of school dropout in Southern Italy. *Giornale degli Economisti e Annali di Economia*, 207-245.

Passaretta, G., Barbieri, P., Wolbers, M. H. J., & Visser, M. (2018). The direct effect of social origin on men's occupational attainment over the early life course: An Italian–Dutch comparison. *Research in Social Stratification and Mobility*, *56*(March), 1–11.

Ponzo, M., & Scoppa, V. (2014). The long-lasting effects of school entry age: Evidence from Italian students. *Journal of Policy Modeling*, *36*(3), 578-599.

Plantenga, J. (2002). Combining work and care in the polder model: an assessment of the Dutch part-time strategy. *Critical Social Policy*, *22*(1), 53-71.

Reardon, S. F. (2008). *Thirteen Ways of Looking at the Black-White Test Score Gap* (IREPP Working Paper No. 2008-08). *IREPP Working Paper*.

Ringarp, J., & Rothland, M. (2010). Is the Grass always Greener? The Effect of the PISA Results on Education Debates in Sweden and Germany. *European Educational Research Journal*, *9*(3).

Ritzen, J. M., van Dommelen, J., & De Vijlder, F. J. (1997). School finance and school choice in the Netherlands. *Economics of education review*, *16*(3), 329-335.

Rözer, J., & van de Werfhorst, H. (2017). *Inequalities in Educational Opportunities by Socioeconomic and Migration Background : A Comparative Assessment Across European Societies*. ISOTIS report.

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*(4), 511–526.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika,* 52(3/4), 591-611.

Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, *75*(3), 417–453.

Skopek, J. (2017). Entry to formal childcare and abilities of preschoolers: A comparison of East and West Germany. In H. Blossfeld, N. Kulic, J. Skopek, & M. Triventi (Eds.), *Childcare, Early Education and Social Inequality – An international Perspective* (pp. 231–248). Celtenham, UK: Edward Elgar.

Skopek, J., & Passaretta, G. (2018). The Social Stratification of Skills from Infancy to Adolescence – Evidence from an Accelerated Longitudinal Design. *SocArXiv Papers*, 1–63. Available at: https://osf.io/preprints/socarxiv/xkctv/

Skopek, J., van de Werfhorst, H. G., Rözer, J., Zachrisson, H. D., & van Huizen, T. (2017). Inequality in Various Stages of the Educational Career: Patterns and Mechanisms – Literature Review. ISOTIS Report, Trinity College Dublin.

Slot, P. L., Leseman, P., Verhagen, J., & Mulder, H. (2015). Associations between structural quality aspects and process quality in Dutch early childhood education and care settings. *Early Childhood Research Quarterly*, *33*, 64-76.

Smeeding T. M., & Rainwater, L. (2004). Comparing Living Standards across Nations: Real Incomes at the Top, the Bottom, and the Middle. In *What Has Happened to the Quality of Life in the Advanced Industrialized Nations?* (pp. 153–183). Luxembourg.

Spiess, C. K., Büchel, F., & Wagner, G. G. (2003). Children's school placement in germany: Does Kindergarten attendance matter? *Early Childhood Research Quarterly*, *18*(2), 255–270.

Statistics Netherlands (2015). Gebruik peuterspeelzaal en formele opvang 2-3 -jarigen naar 20% -groep van het gestandaardiseerde huishoudensinkomen, 2013. The Hague: CBS.

Statistics Netherlands (2016). Kinderen en jaarlijkse uren in formele kinderopvang naar soort opvang en leeftijd van het kind (per einde jaar), 2015. The Hague: CBS.

Steffensen, K., Ekren, R., Zachrisen, O. O., & Kirkebøen, L. J. (2017). *Er det forskjeller i skolers og kommuners bidrag til elevenes læring i grunnskolen? En kvantitativ studie [Are there differences in schools' and municipalities' contributions to student learning in primary school? A quantitative study]*.

Stocké, V. (2007). Explaining educational decision and effects of families' social class position: An empirical test of the Breen-Goldthorpe model of educational attainment. *European Sociological Review, 23*(4), 505–519.

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104*(3), 743-762.

Sullivan, A., Zimdars, A., & Heath, A. (2010). The social structure of the 14-16 curriculum in England. *International Studies in Sociology of Education*, *20*(1), 5–21.

The Norwegian Directorate for Children, Y. a. F. A. (2017). *Oppvekstrapporten 2017 [The report on children's and adolescent's developmental conditions, 2017]*.

Thévenon, O. (2011). Family policies in OECD countries: A comparative analysis. *Population and development review*, *37*(1), 57-87.

UNICEF (2018). *An Unfair Start: Inequality in Children's Education in Rich Countries*. Innocenti Report Card 15. Florence: UNICEF Office of Research – Innocenti.

van de Werfhorst, H. G., & Hofstede, S. (2007). Cultural capital or relative risk aversion? Two mechanisms for educational inequality compared. *British Journal of Sociology*, *58*(3), 391–415.

van de Werfhorst, H. G., & van Tubergen, F. (2007). Ethnicity, schooling, and merit in the Netherlands. *Ethnicities*, *7*(3), 416–444.

van Elk, R., van Der Steeg, M., & Webbink, D. (2011). Does the timing of tracking affect higher education completion? *Economics of Education Review*, *30*(5), 1009-1021.

van Huizen, T., Dumhs, L., & Plantenga, J. (2018). The Costs and Benefits of Investing in Universal Preschool: Evidence from a Spanish Reform. *Child development*, forthcoming.

van Huizen, T., & Plantenga, J. (2018). Do children benefit from universal early childhood education and care? A meta-analysis of evidence from natural experiments. *Economics of Education Review*, *66*, 206-222.

Visser, J. (2002). The first part-time economy in the world: a model to be followed?. *Journal of European Social Policy*, *12*(1), 23-42.

Votruba-Drzal, E., Coley, R. L., Collins, M., & Miller, P. (2015). Center-Based Preschool and School Readiness Skills of Children From Immigrant Families. *Early Education and Development*, *26*(4), 549–573.

West, A. (2006). The pre-school education market in England from 1997: Quality, availability, affordability and equity. *Oxford Review of Education*, *32*(3), 283–301.

West, A., Allmendinger, J., Nikolai, R., & Barham, E. (2010). Decentralisation and educational achievement in Germany and the UK. *Environment and Planning C: Government and Policy*, *28*(3), 450–468.

Winsvold, A., & Guldbrandsen, L. (2009). *Kvalitet og kvantitet - kvalitet i en barnehagesektor i sterk vekst (Quality and quantity - Quality in a child-care sector in rapid growth)*.

Zachrisson, H. D., Dearing, E., Blömeke, S., & Moser, T. (2017). What levels the playing field for socioeconomically disadvantaged children in the Norwegian ECEC model? In H. P. Blossfeld, N. Kulic, J. Skopek, & M. Triventi (Eds.), *Childcare, Early Education and Social Inequality – An International Perspective* (pp. 194–214).

Zeanah, C. H., Boris, N., & Larrieu, J. A. (1997). Infant Development and Developmental Risk : A Review of the Past 10 Years. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(2), 165–178.

Zijsling, D., Keuning, J., M.-A. Keizer-Mittelhaëuser, Naayer, H. &. Timmermans, A. C. (2017). *Cohortonderzoek COOL$^{5–18}$: Technisch rapport meting VO-3 in 2014*. Groningen: GION Onderwijs/Onderzoek.

ISOTIS INCLUSIVE EDUCATION AND SOCIAL SUPPORT TO TACKLE INEQUALITIES IN SOCIETY

This project has received funding from the
European Union's Horizon 2020
research and innovation programme
under grant agreement No. 727069.