RUHR
ECONOMIC PAPERS

Christina Vonnahme

Do Migrant-Native Achievement Gaps Narrow? Evidence Over the School Career

RWI #932

Imprint

Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Editors

Prof. Dr. Thomas K. Bauer
RUB, Department of Economics, Empirical Economics
Phone: +49 (0) 234/3 22 83 41, e-mail: thomas.bauer@rub.de

Prof. Dr. Wolfgang Leininger
Technische Universität Dortmund, Department of Economic and Social Sciences
Economics – Microeconomics
Phone: +49 (0) 231/7 55-3297, e-mail: W.Leininger@tu-dortmund.de

Prof. Dr. Volker Clausen
University of Duisburg-Essen, Department of Economics
International Economics
Phone: +49 (0) 201/1 83-3655, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Ronald Bachmann, Prof. Dr. Manuel Frondel, Prof. Dr. Torsten Schmidt, Prof. Dr. Ansgar Wübker
RWI, Phone: +49 (0) 201/81 49-213, e-mail: presse@rwi-essen.de

Editorial Office

Sabine Weiler
RWI, Phone: +49 (0) 201/81 49-213, e-mail: sabine.weiler@rwi-essen.de

Ruhr Economic Papers #932

Responsible Editor: Thomas Bauer

ISSN 1864-4872 (online) – ISBN 978-3-96973-091-1

Ruhr Economic Papers #932

Christina Vonnahme

Do Migrant-Native Achievement Gaps Narrow? Evidence Over the School Career

RWI

Bibliografische Informationen der Deutschen Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at **http://dnb.dnb.de**

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

http://dx.doi.org/10.4419/96973091
ISSN 1864-4872 (online)
ISBN 978-3-96973-091-1

Christina Vonnahme[1]

Do Migrant-Native Achievement Gaps Narrow? Evidence Over the School Career

Abstract

The integration of foreign origin students in host countries' educational systems has mostly been studied based on cross-sectional data. In contrast, I use data from a national longitudinal education study to calculate achievement gaps in vocabulary, reading and math tests for foreign origin relative to native students over the school career in Germany for the years 2010 to 2018. In line with previous research, the raw gaps are substantial and can be explained to one to two thirds by school characteristics and the socio-economic background of the child. Taking a longitudinal perspective reveals that both raw and conditional gaps slightly decrease over several parts of the school career. However, the unexplained part of the decomposed gaps tends to increase as children grow older. The findings demonstrate that initial disadvantages of foreign origin students reduce rather than accumulate, but partly prevail until the end of school.

JEL-Code: I24, J15

Keywords: Integration; achievement gaps; educational performance; immigrants

December 2021

1 Christina Vonnahme, RWI. – I am grateful to Ronald Bachmann, Thomas Bauer, Julia Bredtmann, Christina Gathmann, Lisa Höckel, Sven Kunze, Fernanda Martínez Flores, Sebastian Otten and Marcus Tamm for their valuable comments and advice. This paper uses data from the National Educational Panel Study (NEPS): Starting Cohorts 2, 3 and 4 (DOIs: 10.5157/NEPS:SC2:8.0.1, 10.5157/NEPS:SC3:9.0.0, 10.5157/NEPS:SC4:10.0.0). From 2008 to 2013, NEPS data was collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network. I thank the team at LIfBi for providing the data and documentation, and for helping with several requests. *– All correspondence to: Christina Vonnahme, RWI, Hohenzollernstr. 1/3, 45128 Essen, Germany, e-mail: christina.vonnahme@rwi-essen.de*

1 Introduction

The size of foreign-born population in OECD countries has been rising continuously in the past decades, with annual growth rates of about 2 percent (OECD, 2020). While international migration offers large benefits to immigrants as well as host country populations, e.g., by improved matching of labor market demand and supply, societies also face challenges caused by rising stocks of immigrants. For labor immigration, obvious barriers to the integration of immigrants emerge when foreign degrees and labor market experience from different countries are not fully recognized. But also for other types of immigration, costs occur, e.g., when immigrants do not speak the host country's language fluently or have to familiarize with a new cultural context. Integration policies at different levels can help to reduce these costs and make immigrants' lives in the host country as pleasant and beneficial as possible, to maximize both natives' and immigrants' welfare.

Also beyond efficiency objectives, the successful integration of immigrants is an important normative goal in many host countries that commit to providing equal opportunities for all members of society, irrespective of their origin, ethnicity or citizenship. Not being discriminated on the basis of these characteristics is a key right for immigrants of all types and ages. However, it can be argued that it is most important for immigrant children, whose successful integration during school can make a difference for their whole life. In this context, a highly relevant research question this paper aims to answer is how the performance of foreign origin students develops over the school career. Knowing whether initial disadvantages are reduced or amplified, is a key prerequisite to the complex assessment of ex post opportunities of foreign origin children (see, e.g., Roemer, 1998; Fleurbaey and Peragine, 2013).

The analyses in this paper add to a large body of literature that has studied the educational performance of immigrant children and the determinants of their integration in school. According to this literature, foreign origin children lag behind their native peers in educational achievement in many countries due to several reasons. These include migration-related disadvantages from language barriers, the lack of knowledge about the educational system, and the often less advantaged socio-economic background of immigrants relative to the native population (OECD, 2016, 2019).[1]

[1]Depending on the type of immigrants a host country attracts, the positive selection of immigrants can

In Germany, the integration of immigrant children in the educational system is particularly relevant. Although public perception had not considered the country a typical "immigration country" for a long time, Germany has been an important host country for international immigrants for several decades. The share of children with foreign origin (i.e., children who have at least one parent without German nationality since birth or a so-called "migration background") amounted to almost 40 percent in 2019 (Federal Statistical Office, 2020).

In international comparison, Germany is also an interesting case to study because migrant-native achievement gaps in education are higher than in most other OECD countries, as shown in several PISA studies (OECD, 2016; Stanat and Christensen, 2006). Over the past waves of this large-scale assessment of student performance, raw achievement gaps in Germany have slightly decreased, while gaps conditional on the socio-economic background increased.

From a methodological point of view, however, it is important to note that comparing results from repeated cross-sections does not allow to draw precise conclusions about whether the educational integration of students has indeed improved. The general problem of changing unobserved characteristics of changing cohorts in cross-sectional studies here applies in a similar way as for the labor market integration of immigrants, where it has been extensively studied since the seminal work of Borjas (1985). Combining data from cross-sectional studies for similar student cohorts observed in different grades can partly address this problem and shed light on the question of the development of foreign origin students' educational achievement. Still, the samples from different cross-sectional studies often do not exactly cover the same student or immigrant cohorts. However, the possibility to analyze the development of achievement gaps over time has been very limited due to the lack of longitudinal education data for many countries.

From a theoretical point of view, gaps can be expected to accumulate over time when initial disadvantages add up, but also to reduce when students with lower initial endowment have higher marginal returns from educational inputs. Previous studies that use panel data have found very different results on the evolution of achievement gaps over time. For the few countries studied so far, ethnicity gaps in education have either widened, stagnated, or even reduced over the school career. Moreover, these studies have mostly been limited to selected years of schooling and are only available for English-speaking, traditional destination countries (see Dustmann et al., 2010

also induce a reverse pattern in certain host countries such as Australia, where immigrant children have better educational outcomes than their native peers (Nguyen et al., 2020; OECD, 2016, 2019).

for the UK, Fryer Jr. and Levitt, 2004 and Bond and Lang, 2018 for the US, and Nguyen et al., 2020 for Australia).

The contribution of the present paper is to fill this gap by providing a comprehensive overview of the migrant-native achievement gap over the school career in Germany, a more recent popular host country in Continental Europe. Data from the National Educational Panel Study (NEPS) for the years 2010 to 2018 provide a unique opportunity to calculate achievement gaps for the same students over a relatively long time period (Kristen et al., 2011) and to investigate their evolution and determinants. The combination of three samples of students enables me to cover the whole educational career, from kindergarten to the end of secondary school. Moreover, the provision of standardized test scores such as vocabulary, reading comprehension, and math allows me to assess the skill development of foreign origin children in the most relevant skill domains, and to compare the results to those of the previous literature.

The determinants of migrant-native achievement gaps may include the migration experience itself and directly related disadvantages of the children or their parents such as transaction costs, language barriers, or discrimination. In addition, the immigrant population often differ from the native host country population in further socio-demographic characteristics that are not directly related to the migration experience, such as the level of education. In countries where 1st-generation immigrants have below average education levels, the gap in educational achievement of their children compared to natives can partly be attributed to this difference, confirming that parental education is a major determinant of children's educational success (OECD, 2018). Next to socio-economic characteristics of foreign origin families, the characteristics of peers and schools can also determine the educational performance of 1st- and 2nd-generation immigrants. Finally, these determinants might affect the choice which school parents send their child to in the first place, which can lead to substantial segregation by origin (as analyzed, e.g., by Hanushek et al., 2009; Schnepf, 2007; SVR, 2013).

To account for differences in socio-economic characteristics, I follow the literature and estimate gaps conditional on proxies for socio-economic background in addition to estimating raw gaps. This procedure has been conducted in previous empirical education studies in line with decomposition analyses for, e.g., labor market integration of immigrants or the gender wage

gap.[2] To control for school-specific determinants, I include school fixed effects that capture time-invariant characteristics of peer, neighborhood, and school quality, as well as further school characteristics such as school track. The longitudinal setting allows me to investigate the development of the relative importance of different types of determinants over time.

The results can be summarized as follows. Migrant-native achievement gaps in vocabulary, reading comprehension and math skills, as measured by standardized tests, are substantial and – with the exception of primary school – the larger the younger the children are. The scope of the raw gaps is on average in line with evidence from previous, mostly cross-sectional studies. For test scores in several skill domains, I find raw gaps of 10 percent and above for most stages of the school career. A slight convergence by 2–5 percentage points between the achievement of foreign origin and native students can be observed for both language and math skills before and after primary school. Conditional gaps are substantially smaller and mostly amount to a maximum of only 5 percent. For some domains and points in time, they are even zero or slightly positive.

Decomposing the raw gap reveals that the importance of school fixed effects or socio-economic characteristics in general decreases over the school career, and the part of the gap that cannot be explained by these characteristics increases as children grow older. This result reflects that conditional gaps increase to a smaller extent than raw gaps over the school career. Further analyses show that achievement gaps differ by school type: raw gaps are smaller for students who make it to the academic track of lower secondary school, but there is less convergence between foreign origin and native students compared to non-academic tracks. For more school-specific measures of achievement, i.e., grades from school reports that form the basis for later school degrees, gaps are smaller in general: they amount to only half the size of the gaps in test scores. Over the course of primary school, the gaps are close to zero, but widen with the start of secondary school and only partly close again in the following years.

Overall, the German school system seems to help foreign origin students to balance initial disadvantages and catch up with their native peers to a certain extent, as achievement gaps in skills decrease over several stages of the school career. However, in some domains foreign origin children still lag behind their native peers even though they have been attending the German

[2]See Section 2 for an overview of the literature on achievement gaps in education and Bartolucci (2014) and Kerm et al. (2016) for decompositions of migrant-native wage gaps. Blau and Kahn (2017) include an overview of the literature on the gender pay gap, and Goldin et al. (2017) emphasize that despite general convergence of the wage differences over time, the gender gap might increase over the individual life.

school system for a long time and even after accounting for family and school characteristics.

The remainder of this paper is structured as follows: Section 2 gives an overview of related literature and the theoretical expectations about the evolution of achievement gaps over the school career. Section 3 presents the institutional background of the educational system in Germany, the data used for the analyses, and the empirical strategy. The results are presented and discussed in Section 4. Section 5 concludes.

2 Related Literature

The performance of students with foreign origin in the educational system of the host country has been scrutinized in a large number of publications. The greater part of this literature relies on data from large-scale national or international education studies, most of which are cross-sectional and repeated on a regular basis. The most prominent study is the Programme for International Student Assessment (PISA) which has been conducted every three years since 2000. It assesses the performance of 15-years-old students in standardized tests in reading, math and sciences, with a rotating focus on one of these domains in each wave. Based on survey information on the students' immigrant background and other socio-economic characteristics, the PISA reports also provide numbers on migrant-native gaps in the respective test domains. The scope of these gaps highly varies by country, but also over time. On average for the participating OECD countries, students with an immigrant background achieve test scores that are 7–10 percent lower compared to their native peers, depending on the skill domain and year. In some countries such as Australia, Singapore or several Arabic countries, raw gaps are positive. When students' or their families' characteristics, such as the socio-eonomic background, are taken into account, the gaps typically shrink to about 3–5 percent. Also for conditional gaps, the range is wide. However, they are positive more often than raw gaps, reflecting the below-average socio-economic status of immigrants compared to natives in many host countries (OECD, 2013, 2016, 2019).

For Germany, raw migrant-native gaps in PISA test scores have been ranging between 10 and 15 percent since 2003 and have been slightly decreasing over time. However still, the gaps are above the OECD average for all domains and waves since 2003. For science test scores in 2015, they have even been second-largest compared to all other countries. Conditional gaps are substantially smaller with 3 percent on average across different domains and waves, but

have been slightly increasing over the years (OECD, 2013, 2016, 2019). The authors of the PISA reports note that the change in the size of the gaps can partly be explained by changing immigrant cohorts. The variables used to calculate conditional gaps include the education level and occupation status of the parents, and the number of books and other educational resources and possessions at home. In selected waves, students' gender, language spoken at home or school characteristics are also accounted for (OECD, 2016).

The PISA data, and data from similar large-scale repeated cross-sectional assessments, have been used in a number of studies to investigate the performance of foreign origin students in more detail. Cross-country analyses that provide an overview of immigrant students' performance in PISA tests include contributions by Entorf and Minoiu (2005), Giannelli and Rapallini (2016) and Stanat and Christensen (2006). While the latter relate the gap to students' school attitudes, motivation and learning strategies, Dustmann et al. (2012) calculate raw and conditional achievement gaps for the 2006 PISA wave and put a focus on Turkish emigrants in different host countries. Further studies combine cross-sectional results from several education data sets. Schnepf (2007), e.g., contrasts the relative performance of immigrant students in ten OECD countries by using data from PISA, TIMSS and PIRLS.[3] Ruhose and Schwerdt (2016) also combine data from these three international studies to analyze the effect of early tracking on achievement gaps.

Some studies that combine different cross-sectional data sets aim to explicitly analyze the development of achievement gaps over time. E.g., Alieva et al. (2018) use a synthetic cohort approach to combine PIRLS, TIMSS and PISA data and show that the migrant-native achievement gap in reading (but not math) narrows when students transit from primary through lower secondary school, i.e., between grades 4 and 9.

Studies that use longitudinal data are rare as most large-scale educational studies are conducted at a cross-sectional level. Most of the longitudinal studies therefore rely on data on a national or regional level. Also here, the results seem to highly depend on the age cohort and the country context. So far, mostly English-speaking traditional host countries have been analyzed, where the focus typically is on ethnicity rather than migrant-native achievement gaps. For the

[3]The Trends in International Mathematics and Science Study (TIMSS) internationally evaluates students' performance in math and science in grades 4 and 8 every four years, while the Progress in International Reading Literacy Study (PIRLS) focuses on grade 4 students and is conducted every five years.

US, Fryer Jr. and Levitt (2004) show that Black-White test score gaps can totally be explained by family characteristics before primary school, but widen by 0.2 standard deviations over the first two years of school, mainly because of school characteristics. By contrast, Bond and Lang (2018) who also use US data, but for a longer period of time, i.e., kindergarten until grade 7, find that the conditional gap between Black and White children does not widen over time when properly accounting for measurement error. Dustmann et al. (2010) calculate achievement gaps in language skills for British students between the age of 6 and 16 and find that ethnicity gaps shrink over the school career, but that heterogeneity between different origin groups is large. For Australia, Nguyen et al. (2020) analyze the evolution of relative academic performance of Asian vs. White students between 4 and 15 years of age. They find that the positive gaps increase along with higher educational time investments of Asian origin families.

The patterns of diverging, stagnating or converging achievements can be explained by the interaction of several effects. On the one hand, it might be the case that initial disadvantages of foreign origin children that arise, e.g., from worse skills in the host country language, add up over time, which would then widen the gaps. This pattern of cumulative effects has been labeled as "skills beget skills" by James Heckman[4] or more generally as "Matthew effects"[5] in the empirical education literature. On the other hand, it could also be that achievement gaps narrow over time when initially disadvantaged individuals harvest higher marginal returns from educational resources, so-called "compensatory effects". Which of the effects prevails in a given context is an empirical question.

Studies that focus on the relative performance of children of immigrants in Germany again mostly rely on cross-sectional data, or data from general household surveys rather than specific education studies. The results are mixed: depending on the specific performance measure and cohort, some studies find negative conditional gaps, while other studies find that controlling for the socio-economic background can fully explain the difference in foreign origin relative to native students' educational outcome (for an overview, see SVR, 2016). E.g. for school attendance and achieved school degrees, Riphahn (2003) finds large raw and conditional gaps, whereas Krause et al. (2015) find no differences for school recommendations and track enrollment between

[4]See The Heckman Equation project (2021) for a synthesis of related literature, and Lubotsky and Kaestner (2016) for a recent empirical analysis on accumulating advantages in kindergarten.

[5]The term Matthew effects refers to a bible quote from the Gospel of Matthew, commonly rephrased as "The rich get richer and the poor get poorer", applied to the sciences context by Merton (1968).

immigrant and native students when accounting for differences in the socio-economic background. Kiss (2013) focuses on grades from school reports and finds conditional gaps in primary, but not secondary education. Ammermueller (2007) decomposes the gap in PISA test scores and shows that family characteristics are particularly important at the lower part of the test score distribution, but that parental preferences and time of enrollment are even more decisive in general. For the NEPS data, Olczyk et al. (2016) calculate raw gaps for different groups of immigrants in the first wave of the study.

Overall, the literature has identified migrant-native achievement gaps from cross-sectional data that are negative for many countries in Continental Europe including Germany, but much smaller or sometimes zero when controlling for the socio-economic background of the students. In traditional, mostly English-speaking, immigration countries, the conditional gaps are usually smaller. While most evidence is generated from cross-sectional data, the few longitudinal studies find converging, stagnating as well as diverging gaps over the school career. The evolution of migrant-native achievement gaps thus seems to be highly dependent on the context of the host country and its educational system as well as on the characteristics of the immigrants and their children.

For Germany, the development of achievement gaps between foreign origin students and their native peers over the school career has not been investigated with large-scale nationally representative panel data yet. I contribute to the literature by estimating migrant-native achievement gaps over the school career based on NEPS data. This longitudinal study has been conducting standardized tests in different skill domains for several cohorts of children since 2010. It covers the range from kindergarten to the end of secondary school and thus provides an outstanding opportunity to shed light on the relative performance of foreign origin students over several parts of the school career.

3 Institutional Background, Data, and Empirical Strategy

3.1 The Educational System in Germany

I now briefly outline the general setup of the early education and school system in Germany as a background for the following analyses of migrant-native achievement at different stages of the school career. A summary of the types of educational institutions is provided in Figure 1.

In early childhood, children in Germany are eligible to public daycare facilities from the age of one year on.[6] Parents can choose between different time schedules for public daycare. The costs are in general income-based and vary on the community level. In some federal states or communities, parents do not have to contribute to the costs at all, while in others, parents with high income must pay several hundred euros per child and month. Over the past years, these costs have been substantially reduced.[7] The use of public childcare is voluntary and increases with children's age.[8] It typically depends on the socio-economic background of the family. Children from households with lower income and lower parental education as well as children with foreign origin are less likely to visit public daycare. However, these differences balance toward the end of kindergarten, when 90 percent of foreign origin and 97 percent of native children attend public daycare (Autorengruppe Bildungsberichterstattung, 2018). Accordingly, when calculating achievement gaps for kindergarten ages, it is important to be aware that attendance can be selective and the selection might differ between foreign origin and native children.

Primary school is mandatory for children from the age of about six years.[9] It normally offers lessons for about four to five hours per day. In the afternoon, many primary schools offer voluntary additional care that partly requires financial contributions by the parents. Here, children are supervised when doing homework or extracurricular activities. In 14 of the 16 German federal states, primary school lasts for four years. Only in Berlin and Brandenburg, it includes grades 5 and 6 as well.[10] Primary school is thus the stage of the school career where

[6] Before August 2013, the age threshold for a legal claim had been higher at three years.

[7] Busse and Gathmann (2020) and Cornelissen et al. (2018) evaluate the effects of public childcare on maternal labor supply and child outcomes.

[8] In 2019, one third of children under the age of three was in public childcare, and 93 percent of children between the ages 3–6 (Autorengruppe Bildungsberichterstattung, 2020). For children aged 4–5 years, the numbers were even higher with up to 98 percent (for the year 2017, Autorengruppe Bildungsberichterstattung, 2018).

[9] The exact timing varies on the federal state level, as do many further regulations and peculiarities of the educational system in Germany.

[10] After primary school, some federal states offer joint learning for 'orientation' in grades 5 and 6 before children or their parents choose a specific track. In the NEPS setting, these are considered in the same way as federal states without an 'orientation phase'.

only little selection can be expected. In most federal states, also regional selection is limited and children have to attend a primary school close to where they live. Accordingly, selection issues here can mostly be attributed to residential choice.

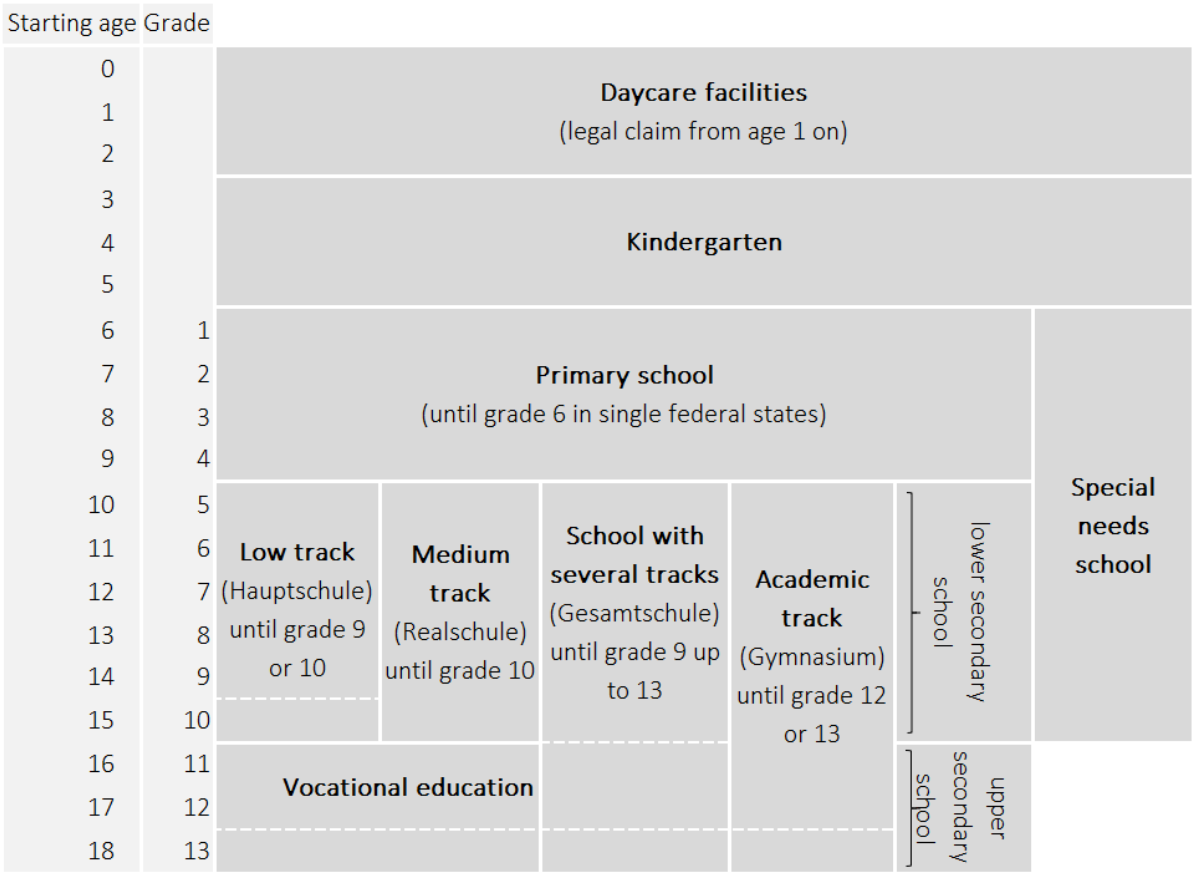

Figure 1: Overview of the Educational System in Germany

Source: Author's illustration.

At the end of primary school, when children are approximately ten years old, teachers give recommendations for which track each child should attend thereafter. These recommendations are relatively binding in some federal states, but can be completely ignored in others. The different tracks to choose from include an academic track ("Gymnasium"), a medium track ("Realschule") and a low track ("Hauptschule"). In the academic track, children can directly proceed from lower to upper secondary education after grade 10 which marks the end of compulsory schooling. An upper secondary degree (A-levels) is required for tertiary education. The medium and the low track have a stronger focus on teaching applied skills and preparing the students for vocational training. A good degree after grade 10 from the medium or low track qualifies switching to the academic track. Also during lower secondary education, transitions between the different tracks

are possible, but not very common.[11] In some federal states, schools that offer the low track have been abolished in the past years (or, in the case of states in Eastern Germany, never existed in separate schools) or have been joined with the medium track. In addition to schools for three different tracks, comprehensive schools ("Gesamtschulen") exist in some states. They offer all tracks separately or award different degrees to the high- vs. low-performing students within single classes, but always also offer upper secondary schooling equivalent to the academic track. Further types of free schools (e.g., "Waldorfschulen") are most similar to comprehensive schools.

After grade 10, students who do not continue with the academic track can directly transition to vocational training that either takes place in schools only or in the dual vocational system. Upper secondary education lasts two to three years, depending on the federal state and the degree aimed at. In general, selection based on socio-demographic characteristics including foreign origin is strongest for lower and upper secondary school. The calculation of achievement gaps should thus take into account different track types.

3.2 Data: National Educational Panel Study

Description of Data Set and Sampling

The National Educational Panel Study (NEPS) is the first large and representative longitudinal education survey in Germany. It has been conducted on an annual basis since 2010 and thus allows following individuals along their education careers for a long period of time.[12] The NEPS combines surveys of children, their parents, educators and school principals with own standardized tests that allow analyzing the development of competencies over time. The study follows a cohort design. It includes six so-called "starting cohorts" (SCs), i.e., samples of individuals from different birth cohorts and stages of their education career. I focus on the second, third and fourth starting cohort, as these cohorts cover children over their school career. SC2 (which I call "child cohort" hereafter) covers children who attended the penultimate year of public daycare in 2010/2011 and enrolled at primary school in 2012. They reach the second grade of lower secondary school (grade 6) in the school year 2017/2018. SC3 (which I call "teen cohort") includes children from

[11] In the school year 2018/19, 2.5 percent of all students from grades 7–9 changed to a different school, out of which 5.6 percent changed to the academic track and 52.5 percent left the academic track (Autorengruppe Bildungsberichterstattung, 2020).

[12] In some years, even two waves were realized to target specific sub-samples or measure specific items.

their first year at secondary school, i.e., grade 5 in the school year 2010/2011. This cohort is followed until upper secondary school (grade 12 in the school year 2017/2018). SC4 (which I label "adolescent cohort") starts in grade 9 in the school year 2010/2011 and follows students through the end of lower secondary, the whole upper secondary school and thereafter. Table 1 provides an overview of the age ranges and grades covered by the three starting cohorts.[13] The overall setup of the data set is described in Blossfeld et al. (2011).

Table 1: Overview of Starting Cohorts (SCs)

SC (cohort no.)	Starting grade	Average age in starting grade	Final grade	Average age in final grade
2 "child"	penultimate kindergarten year	4–5	grade 6 (lower sec. school)	11–12
3 "teen"	grade 5 (lower sec. school)	10–11	grade 12 (upper sec. school)	17–18
4 "adolescent"	grade 9 (lower sec. school)	14–15	grade 12/13 (end upper sec. school)	18/19+

Source: Author's illustration.

The sampling of the three cohorts I use is implemented as to make each starting cohort representative for its respective target population. The sampling procedure follows an implicit stratification with respect to federal state, regional classification and funding of the institution (as some kindergartens or schools are at least partly privately funded, e.g., when run by churches). Participation in the study is voluntary for both educational institutions as well as individual participants (i.e., children and their parents). Kindergartens or schools who refused to participate were replaced by institutions from the same strata. For SC2 starting in kindergarten, an indirect sampling design was chosen that first sampled primary schools and then kindergartens from lists of each primary school's typical "supplier kindergarten". Children who attended other primary schools than those originally sampled were individually tracked, but not tested any more in most of the subsequent waves. With the transition to primary school, the sample of the child cohort was augmented by peers who had attended other than the "supplier kindergartens", and by primary schools where none of the students had participated in the survey during kindergarten. As a result, the sample of the child cohort is relatively small for the kindergarten years and experiences a large augmentation in the first year of primary school. Similarly, the teen cohort experienced a sample augmentation after two years because secondary education (in the tracked system) only starts in grade 7 in two federal states. Within each sampled school, up to two classes

[13]Further starting cohorts, which I do not use in my analysis, include a sample of newborn (SC1), university students (SC5) and an adult cohort (SC6).

were randomly selected. Students who leave the institution they were originally sampled in are tracked and surveyed individually. Details of the sampling procedure and the resulting samples for the three cohorts are provided by Steinhauer et al. (2016), Steinhauer and Zinn (2016a), and Steinhauer and Zinn (2016b), respectively. Aßmann et al. (2011) describe the general sampling procedure for all cohorts.

Description of Outcome Variables

The NEPS data provide different measures of educational achievement. These include test scores in different skill domains as well as grades from annual (or sometimes even half-yearly) reports. The standardized achievement tests are constructed by the NEPS team and conducted by the interviewers. Thus, they are comparable across classes and schools, but are not available every year as opposed to school report grades. School grades, in turn, are more subjective or context-dependent measures of educational achievement as they are awarded by the teacher and mostly not based on central exams.

For the main analyses, I start using test scores as most objective measures of educational achievement. All test scores aim to measure important skills that the tested individuals acquire over their educational career. The NEPS provides test scores for several skill domains such as vocabulary, reading, grammar, orthography, math, sciences, ICT skills, and English (see Weinert et al., 2011 for an overview of the test scores in different domains and cohorts conducted in the NEPS). As each test takes around 30 minutes to complete, time restrictions prevent that each domain can be tested each year. To analyze the development of foreign origin children's relative to native children's achievement over time, I focus on those test scores that are available at several points in time for most of the cohorts, i.e., tests in vocabulary, reading competence, and math. The domains of host country language and math have typically been used in previous studies on migrant-native achievement gaps and hence analyzing these domains facilitate comparability of the results. Also, these tests are the most general proxies for achievement in the two broader skill domains of language and science. Analyzing both of these domains when comparing the achievement of foreign origin vs. native students is particularly relevant to investigate whether foreign origin students lag behind more in language-related subjects, or if gaps also exist for

areas where migration-related disadvantages are less obvious.[14]

For language skills, several sub-domains are tested (see Berendes et al. 2013 for an overview). A vocabulary test is conducted at three points in time for children of the child cohort (SC2) and once for teens and adolescents (SC3 and SC4). Thus, the vocabulary test can be used to measure the development of language skills for young children. Tests of reading competence are conducted every 2–3 years from grade 4 on. While only one test is available for the child cohort, for students from the teen (adolescent) cohort 3 (2) test results are available, which again allows for a longitudinal analysis of skill development. The tests are similar to those used in other educational studies such as PISA.

The vocabulary test is based on the Peabody Picture Vocabulary Test (PPVT, see Dunn and Dunn, 1981), which is often used as a reference for tests in other educational studies. It comprises between 66 and 89 items, depending on the cohort and wave. Here, target words are read aloud to the students, who then have to choose the matching picture out of four pictures presented. The vocabulary test can be seen as a proxy for listening comprehension, accumulated knowledge and crystallized intelligence (Berendes et al., 2013).

The reading competence test mostly consists of multiple-choice questions, which are presented for five different text types with the following functions: information texts, texts with a commenting or argumenting function, texts with a literary-aesthetic function and advertising texts. For each type of text, three different types of comprehension or cognitive requirements are tested. They include finding information, drawing text-related conclusions and reflecting/assessing. The test consists of 31–46 items. Gehrer et al. (2013) describe the framework for assessing reading competence in detail.

Math competencies are tested most regularly, at four points in time for each of the child and teen cohort and two waves during school for the adolescent cohort.[15] The tests cover the four content areas Quantity, Change and relationships, Space and shape, and Data. At the same time, the six cognitive processes Mathematical communication, Mathematical argumentation, Modeling, Using representational forms, Mathematical problem solving, and Technical abilities

[14]Previous literature has shown that language skills are an important prerequisite for performance in math tests, too, as learning and doing exams in math is more difficult for non-native speaking students (Soto-Calvo and Sánchez-Barrioluengo, 2016; Isphording et al., 2016).

[15]I only cover the most relevant time period and exclude the first test score for the child cohort in kindergarten and the last test score for the teen cohort at the end of upper secondary schooling, as including these waves would substantially reduce the balanced panel I construct, as described in the following section.

and skills are tested. In total, 21–33 items are included, mostly administered as multiple choice questions. They are similar to the items used in the PISA study (for 9^{th}-graders) or the German Mathematics Education standards (for 4^{th}-graders). For all other ages, tests that suit the skill development at the respective age are developed. Neumann et al. (2013) describe the tests on mathematical competencies in detail.

The scaling of the test scores is done in accordance with Item Response Theory (IRT), which allows calculating a common scale for individuals' performance in these tests at different points in time. This standardization is facilitated by the use of a linkage design that relies on items that are asked in several waves, if applicable. Most items vary across waves as to reflect students' development in competencies. Even though individuals receive age-specific tests as they grow older, the common scale allows the analysis of the development of the tested competency over time. In the NEPS, the scaling of the test scores occurs either via sum scores (i.e., the number of correct items, for vocabulary) or weighted maximum likelihood estimates (WLE, for reading and math) that ensure unbiased estimates of the latent competencies by properly considering different types of missing values and test fatigue effects that arise when several tests are conducted after one another. The WLE scores have a mean of zero in the first wave and a standard deviation of about one. The values in subsequent waves may vary around the initial values (Pohl and Carstensen, 2012, 2013; Fischer et al., 2016).

Grades in German and math from students' annual reports are available each year from grade 3 on, for some grades even two times per year (from half-yearly reports). I use them as additional outcomes. They are measured on a scale from 1 (very good) to 6 (insufficient). Grade 4 is "sufficient" to pass a subject. In upper secondary schooling, an alternative reversed scale from 0–15 points is applied, where 0 corresponds to insufficient and 13–15 points to "very good". This point system allows for a more precise measurement of grades after lower secondary schooling. As grades are granted by teachers based on students' performance during the school year, e.g., in exams and other types of participation in the class, they are more subjective measures than objectively measured test scores. Except for exams for certain school-leaving degrees, exams are typically designed and graded by the teacher. The demands here depend to a certain degree on the specific school and class.

I re-scale all outcome variables to having a mean of 100 for native children and a minimum of zero in each cohort-specific wave. This linear transformation leaves the relative size of the gaps unaffected, but ensures that the gaps of foreign origin children are measured relative to natives' achievement at each point in time and that the difference can be interpreted directly as percent deviations. As a result, gaps can easily be compared across outcomes and to results from other studies.

Sample Restrictions and Description

The sample used for the main analyses consists of children and youth from the three described cohorts. To avoid bias from selective attrition and non-response and to analyze a long time period, I construct a balanced panel by conditioning on non-missing values for the outcome as well as main control variables in all of the waves used. It should be noted, though, that conditioning on vocabulary test scores in kindergarten comes at the cost of substantially reducing the sample size for the child cohort, and conditioning on outcomes in grades 5 and 6 means to exclude the states of Berlin and Brandenburg for the teen cohort. Constructing a balanced panel for the adolescent cohort from grade 9 to grade 12 has the effect that most of the students included here attend the academic track in a "Gymnasium" or in a school with several tracks ("Gesamtschule"). To keep the sample as representative as possible, I impute missing values for the school or kindergarten ID, as well as individual and family covariates from adjacent years, and I introduce missing categories for family characteristics which allows me to include children whose parents never took part in the survey throughout the panel to minimize sample selectivity. This is the case for one third of the children from the gross sample (on average for the three cohorts). I exclude students from special needs schools as test scores are only available for a very small number of students attending these schools.

I abstain from using weights for two reasons. The weights provided with the data set are not consistent across waves and cohorts, and longitudinal weights are only provided for continuous participation – even in waves without the test scores I use. Thus, using these weights would come at the cost of reducing the sample and introducing further selection bias. However, unweighted estimates should be unbiased as well given that the sampling procedure already aimed at providing a representative data set for the respective target population of each starting cohort.

Table 2: Summary Statistics

	Native		Foreign origin		
	Mean	SD	Mean	SD	Difference
Test scores and grades					
Vocabulary	100.00	17.49	84.88	21.91	15.12***
Reading competence	100.00	27.68	90.38	28.91	9.62***
Math	100.00	25.16	90.10	26.30	9.90***
German grade	100.00	24.76	95.64	24.99	4.36***
Math grade	100.00	28.76	95.16	29.89	4.84***
Individual and family characteristics					
Female	0.51	0.50	0.54	0.50	−0.02***
Age	13.15	3.35	13.34	3.29	−0.19***
Born abroad	0.01	0.09	0.15	0.36	−0.14***
German citizen	1.00	0.00	0.83	0.38	0.17***
Number of books at home					
0 to 25	0.09	0.28	0.23	0.42	−0.15***
26 to 100	0.20	0.40	0.28	0.45	−0.08***
101 to 200	0.22	0.41	0.20	0.40	0.02***
201 to 500	0.28	0.45	0.17	0.38	0.11***
>500	0.22	0.41	0.12	0.32	0.10***
Mother's education level					
Low	0.02	0.15	0.18	0.39	−0.16***
Medium	0.57	0.50	0.38	0.49	0.19***
High	0.21	0.41	0.09	0.28	0.12***
Missing information	0.20	0.40	0.35	0.48	−0.15***
Father's education level					
Low	0.01	0.09	0.13	0.34	−0.12***
Medium	0.40	0.49	0.29	0.45	0.11***
High	0.24	0.43	0.11	0.31	0.13***
Missing information	0.35	0.48	0.47	0.50	−0.12***
Mother's occupation status					
Blue-collar	0.06	0.24	0.12	0.33	−0.06***
White-collar low-skilled	0.27	0.44	0.19	0.39	0.08***
White-collar high-skilled	0.44	0.50	0.28	0.45	0.16***
Missing information	0.23	0.42	0.41	0.49	−0.18***
Father's occupation status					
Blue-collar low-skilled	0.07	0.25	0.12	0.32	−0.05***
Blue-collar high-skilled	0.09	0.29	0.10	0.29	−0.00
White-collar low-skilled	0.08	0.27	0.06	0.24	0.02***
White-collar high-skilled	0.37	0.48	0.21	0.41	0.15***
Missing information	0.39	0.49	0.51	0.50	−0.11***
Type of educational institution					
Kindergarten	0.01	0.08	0.00	0.06	0.00***
Primary school	0.25	0.43	0.23	0.42	0.02***
Low track (Hauptschule)	0.06	0.24	0.13	0.33	−0.07***
Medium track (Realschule)	0.15	0.35	0.16	0.36	−0.01***
Academic track (Gymnasium)	0.47	0.50	0.39	0.49	0.08***
Several tracks (Gesamtschule, Fachschule)	0.07	0.26	0.10	0.29	−0.02***
Observations	59,037		16,228		75,265

Notes: The table shows summary statistics for students from all three starting cohorts. Values for test scores and grades are conditional on students in the balanced panel for the respective outcome, i.e., who are observed at each point in time and included in the main analyses. Values for further variables are shown for individuals from sub-samples for any outcome. Columns (1) and (2) show values for native students and columns (3) and (4) show values for students of foreign origin, i.e, of whom at least one parent was born abroad. The last column shows the difference in means for the two sub-samples. – Asterisks indicate p-values according to: *** p<0.01, ** p<0.05, * p<0.1. – *Source*: Own calculations based on NEPS data, Starting Cohorts 2–4, 2010–2018.

Table 2 shows summary statistics for the overall sample used in the main analyses. As test scores and grades are not measured in each cohort and wave, the numbers for these outcome variables refer to the balanced panel by outcome. For the remaining variables, numbers for the overall sample are shown: each observation included here may appear in sub-samples for several outcomes. Columns (1) and (2) show the mean and standard deviation for native children. Columns (3) and (4) refer to children with foreign origin. I define a child to have a foreign origin if at least one parent was born abroad. In turn, children are considered native if their parents were born in Germany. The comparison of Columns (1) and (3) shows the extent of raw achievement gaps averaged across all waves. The gap amounts to more than 15 percent for the vocabulary test. For reading comprehension and math, the gap almost reaches 10 percent. For grades in math and German, it is considerably lower with just below 5 percent. Figures A1 and A2 show the full distribution of the achievement variables for native vs. foreign origin students.

Native and foreign origin children also differ with respect to individual and family characteristics. Foreign origin children are slightly older and the sample includes more girls than the one for native children. 15 percent of foreign origin students were born abroad, and 83 percent are German citizens. Foreign origin children on average have a less advantageous family background in terms of the number of books at home and direct measures of parental education and occupational status. As the outcomes are mostly only tested from primary school on, the share of children in the balanced sample who attend public childcare ("Kindergarten") is only about 1 percent. 25 percent of the observations are in primary school. Thereafter, significantly less native children attend the low track or schools with several tracks, but rather attend the academic track. The share of native and foreign origin children in the medium track is similar.

3.3 Empirical Strategy

I follow the literature and calculate migrant-native achievement gaps for different points in time of the school career by estimating the following equation:

$$y_{icws} = \alpha_{cw} + migrant_i'\beta_{cw} + \gamma_s + X_{icws}'\theta + \epsilon_{icws} \quad (1)$$

Here, y_{icws} describes an educational outcome such as a test score or grade of child/student i in cohort-specific wave cw in kindergarten or school s. α_{cw} captures cohort-wave-specific intercepts

for natives[16]. These waves include the penultimate year of kindergarten to grade 6 of lower secondary school for the child cohort, grade 5 to grade 10 for the teen cohort, and grade 9 to grade 12 for the adolescent cohort. The main coefficients of interest are included in the vector β_{cw}, which measures the average gap of foreign origin children's relative to natives' achievement in each cohort-specific wave. The dummy $migrant_i$ is equal to one for foreign origin students, i.e., children or youth of which at least one parent was born abroad. γ_s describes school or kindergarten fixed effects that are added when estimating conditional gaps. In a second step, I also control for socio-economic characteristics of the individual i and their family, X_{icws}, the effects of which are included in θ. ϵ_{icws} describes the error term. I cluster all standard errors at the school (or kindergarten) level.

The choice of control variables to estimate conditional gaps follows previous literature, e.g., cross-sectional education studies such as PISA (OECD, 2016). Here, variables to capture the socio-economic background of the family typically include education levels and occupation types of the parents, as well as home possessions such as books. Partly, also certain school characteristics are accounted for when calculating conditional achievement gaps. To facilitate the comparability with results from previous studies, I closely follow OECD (2016) and include in X_{icws} 5 categories for the number of books at home and 4–5 categories for each of the mother's and father's education level and occupation status (see Table 2 for a full description of the control variables). Moreover, a dummy for the child's gender is included. The school fixed effects γ_s capture time-invariant characteristics of peer, neighborhood and school quality as well as further school characteristics such as school track.

The difference between the raw and the conditional gap is equivalent to the explained part of a twofold Kitagawa-Oaxaca-Blinder decomposition, as described in Fortin et al. (2011).[17] I further decompose the explained part into the share that can be attributed to school fixed effects vs. socio-economic background characteristics, to investigate the relative importance of different determinants and how the different shares develop over the course of the school career.

[16]In the specifications without further covariates, they all amount to 100 due to the common scaling of the outcome variable.

[17]The decomposition approach developed by Kitagawa (1955) was applied to the labor market setting by Oaxaca (1973) and Blinder (1973).

4 Results

4.1 Main Results

The gaps in educational achievement of foreign origin relative to native students over the school career are depicted in Figure 2 for test scores in different skill domains: Vocabulary (Panel A), Reading comprehension (Panel B) and Math (Panel C). Each panel shows results for the child, teen, and adolescent cohort, divided by vertical lines in the graphs. As discussed in Section 3, different domains are tested at different points in time, such that the number of estimates varies across cohorts and test domains. Accordingly, also the number of observations differs between the different domains. All graphs show achievement gaps estimated as specified in Equation 1, in pairs of raw and conditional gaps. As the dependent variables are scaled to a mean of 100 for natives and a minimum of 0, the raw gaps can be interpreted as percent deviations in mean differences for foreign origin vs. native students at the specified grade. The second coefficient of each pair then adds kindergarten or school fixed effects (labeled "Institutional FE", which capture peer and neighborhood quality as well as time-invariant characteristics such as school track and quality) and also adds individual and family characteristics, as described in Section 3.3. Tables A1 to A5 include the estimates behind the graphs discussed, along with additional statistics.

Panel A of Figure 2 shows that the raw gap, i.e., the mean difference in vocabulary test scores for foreign origin vs. native children amounts to 25 percent in the penultimate year of kindergarten ("K1") for the child cohort (SC2). When institutional fixed effects as well as individual and family characteristics are controlled for, the gap is more than halved and shrinks to just below 12.5 percent. When children proceed to primary school ("G1" and "G3" denote grade 1 and grade 3, respectively), the raw gap almost reduces by 4 percentage points, but then does not decrease further during primary school. Rather, it slightly widens again by 2 percentage points in grade 3. The conditional gap that accounts for average disadvantages of foreign origin students with regard to socio-economic characteristics decreases even more when children transition from kindergarten to primary school. It then amounts to 6–8 percent, i.e., about one third of the raw gap and about half the size of the conditional gap in kindergarten. For the sub-samples from the teen and adolescent cohort (SC3 and SC4 – here, students in grade 6 and 9, the raw gaps are smaller with only 16 and 14 percent. The conditional gaps are again up to half the size of

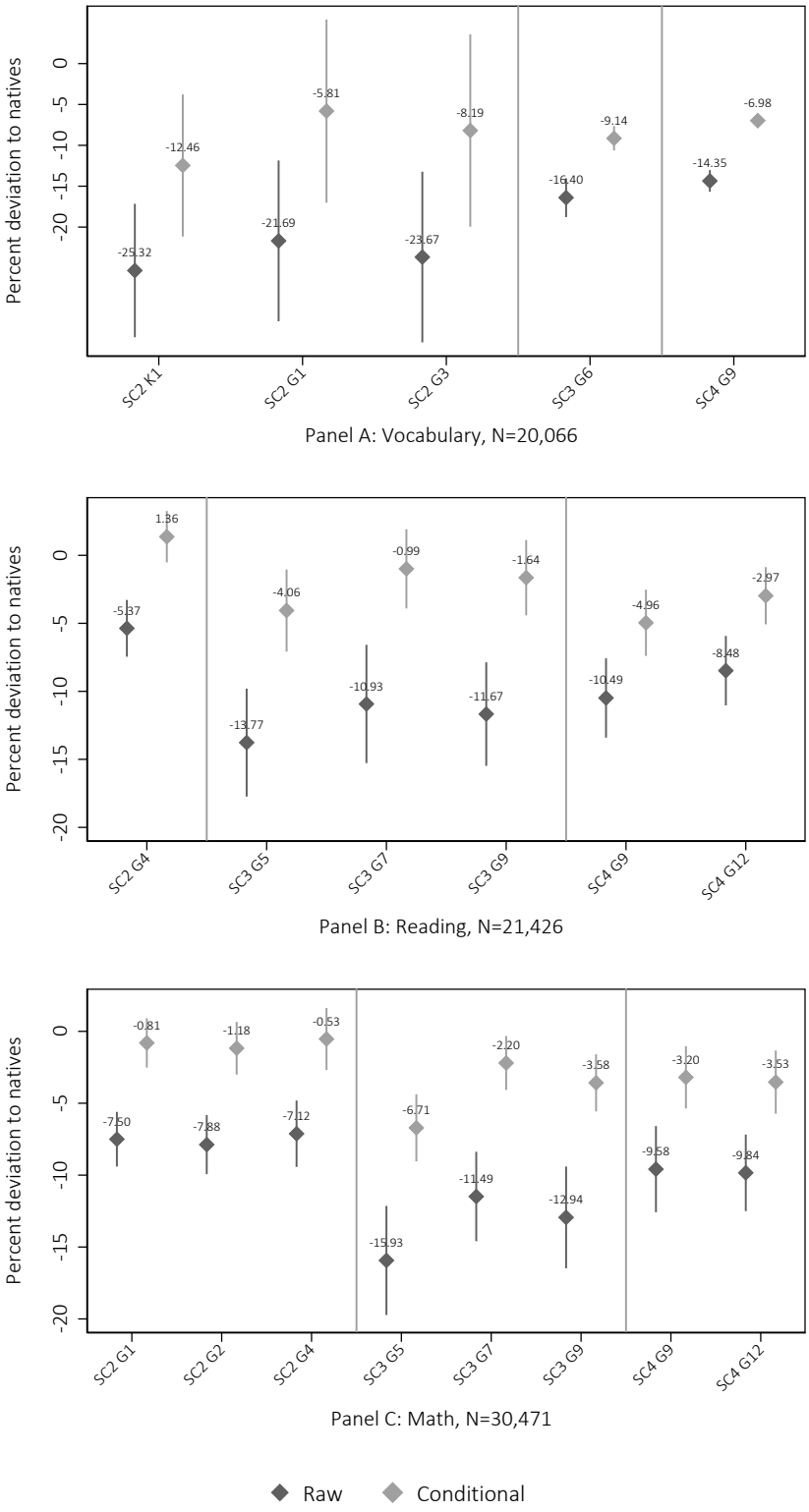

Figure 2: Migrant-Native Achievement Gaps for Test Scores

Notes: Coefficient estimates for raw and conditional achievement gaps of foreign origin (i.e., where at least one parent was born abroad) relative to native students for different starting cohorts (SC2–4) and grades (from "K1", the penultimate year of kindergarten for SC2 to "G12", i.e., grade 12, for SC4). Conditional gaps are estimated by controlling for kindergarten or school fixed effects and individual and family characteristics as described in Equation 1 and presented in Tables A1 to A3. Table 2 includes a full description of control variables. Vertical bars depict 95% CIs. – *Source*: Own calculations based on NEPS data, Starting Cohorts 2–4, 2010–2018.

the raw gaps. Overall, the pattern of achievement gaps over the early educational period shows some convergence. However, the differences between the gaps are statistically insignificant at the 95 percent level, as the corresponding confidence intervals largely overlap. Standard errors, all of which are clustered at the kindergarten or school level, are particularly large for the child cohort as the number of children who are followed from kindergarten to primary school is small (see Section 3).

Panel B of Figure 2 shows corresponding achievement gaps for reading test scores. Raw gaps are relatively small with 5 percent for primary school students in grade 4 and even turn slightly positive once school fixed effects and socio-economic characteristics are added. For older students from the teen cohort (SC3), both raw and conditional gaps are larger, starting at 14 and 4 percent in grade 5. When students proceed through lower secondary school, the raw gap reduces to 10.9 percent in grade 7 and slightly increases again thereafter (11.7 percent in grade 9). The conditional gap amounts to only 1 percent in grade 7 and slightly widens to 1.6 percent in grade 9. For foreign origin students at the middle and end of lower secondary schooling, most of the difference in reading test scores compared to their native peers can thus be explained by the schools they attend and their disadvantageous socio-economic background. For students from the adolescent cohort (SC4) who transition from lower to upper secondary school, i.e., who mostly attend some type of academic track after grade 10, the raw gaps are slightly smaller compared to students from the same grade from the teen cohort and decrease by 2 percentage points until the end of upper secondary schooling (grade 12). The size of the conditional gaps is again about half the size of the raw gaps, i.e., 5 and 3 percent. Here, some convergence in achievement within a cohort can be observed for the transition from lower to upper secondary school, similar to what was observed for the transition from kindergarten to primary school for vocabulary test scores. While the school fixed effects and socio-economic background of foreign origin students explains most of the gap relative to native students in grades 7 and 9 for the teen cohort (SC3), substantial gaps of 3–5 percent remain for students from the adolescent cohort (SC4) that cannot be explained by the observed characteristics.

For math skills, Panel C of Figure 2 draws a picture similar to the one for reading skills. The main difference is that both raw and conditional gaps are slightly more negative (i.e., larger in absolute terms) than gaps for reading skills, which underlines that disadvantages of foreign

origin children in the educational system must not be limited to language skills, but can translate to other subjects. As opposed to reading skills, the math skills of the youngest children (from SC2) are not only tested once, but at three different points in time, which allows studying the development of skills between grades 1 and 4. Both the raw and the conditional achievement of foreign origin relative to native students' achievement merely vary over the course of primary school. The raw difference between the groups' test scores amounts to 7–8 percent. The difference shrinks to 0.5–1.2 percent when taking the school fixed effects and the socio-economic background of the students into account. For older students from the teen cohort (SC3), raw and conditional gaps are also slightly larger than for reading. Raw gaps here start at 16 percentage points in grade 5 and then decrease to 11 percent in grade 7. Towards the end of lower secondary schooling in grade 9, they slightly increase again by 1.5 percentage points. Controlling for school fixed effects and socio-economic background characteristics more than halves the gap in grade 5 and even reduces it to 2.2–3.6 percent in later grades. For upper secondary schooling, the raw gaps are smaller than for the teen cohort with only 9–10 percent, but the conditional gaps are comparable in size with 3.2–3.5 percent.

Overall, some convergence in raw as well as conditional achievement of foreign origin relative to native students can be observed over the school career for different skill domains. Gaps in vocabulary are largest and slightly decrease between kindergarten and primary school. However, comparing the gaps to those of older cohorts also shows that the raw gaps remain large over the rest of the school career and can only be explained by one to two thirds by observable characteristics in secondary school. For reading comprehension, the raw gaps are small in primary school, but wider for the teen cohort at the beginning of lower secondary school. Towards the end of lower secondary education they become smaller again, i.e., a certain decrease in raw and conditional gaps can be observed here also. For the adolescent cohort (SC4, observed from lower to upper secondary education), where most students attend the academic track, raw gaps are again smaller compared to the the teen cohort (SC3), but conditional gaps are larger. This suggests that differences may also exist with respect to school track. For math, the raw and conditional gaps merely change over primary school, but conditional gaps are relatively small here. For secondary schooling, the gaps are larger again (similar to the pattern for reading test

scores), and also here, a certain convergence of foreign origin and native students' achievement over secondary school can be observed.

The pattern of a slight decrease in achievement gaps over the school career raises the question which factors explain the partial convergence in achievement and the difference between the raw and conditional gaps. To answer this question, Figure 3 plots the unexplained part of the gap, i.e., the share of the conditional relative to the raw gap, along with the shares of the raw gap that can be explained by school fixed effects (labeled "institutional characteristics") versus individual and family characteristics. For vocabulary test scores (see Panel A), the figure shows that kindergarten fixed effects only explain a very small part of the achievement gap between foreign origin and native students. About half of the gap is explained by individual and family characteristics, and almost half stays unexplained. Which primary school a child attends matters more than the specific kindergarten, and the specific school is even more important for the cohorts in secondary education. However, the unexplained part is also higher for secondary than for primary school, while the importance of individual and family characteristics decreases over time.

For reading skills (Panel B), the unexplained part is negative for the youngest cohort in primary school, which reflects that a positive conditional gap is observed. Because conditional gaps are relatively small in lower secondary education for the teen cohort (SC3), the unexplained part is small here. As for gaps in vocabulary in secondary school, the importance of individual and family characteristics is relatively limited for reading skills when children grow older. Again, the unexplained part is larger for the oldest, the adolescent cohort (SC4). Also for math skills (Panel C), the picture is similar: individual and family characteristics explain a large part of the gap in primary school, but lose importance in lower secondary school for the teen cohort. There, the specific school attended is most important. For the oldest cohort that transitions from lower to upper secondary schooling, the unexplained part is higher again, and individual and family characteristics also re-gain importance.[18]

The fact that a substantial part of the achievement gaps cannot be explained by observable characteristics of the families or schools shows that the educational system is not able to fully balance initial disadvantages of immigrant children over the school career. This pattern may

[18]Recall that most of the adolescent cohort attends the academic track as the balanced panel only includes students who proceed until upper secondary school.

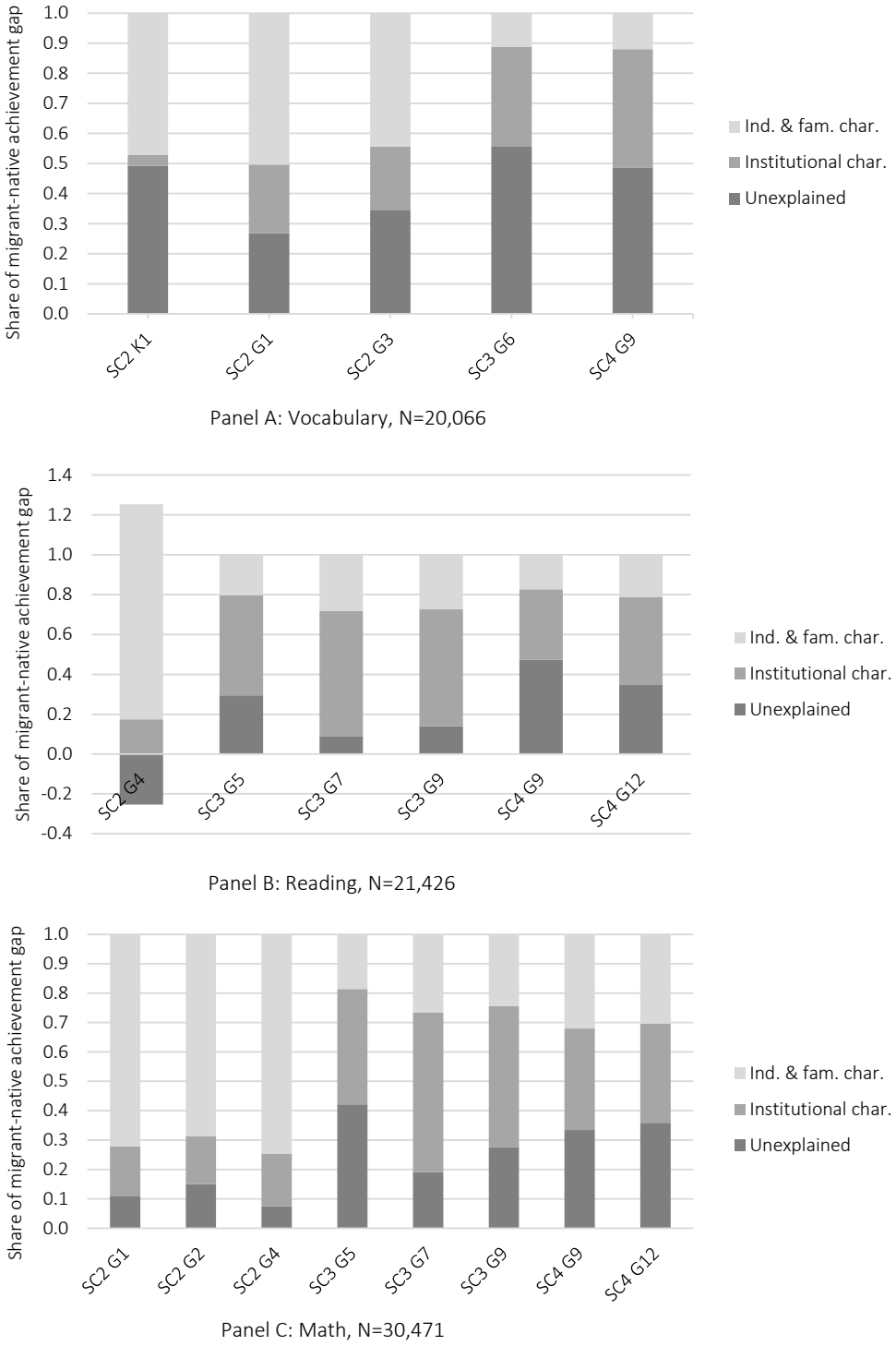

Figure 3: Decomposition of Migrant-Native Achievement Gaps for Test Scores

Notes: The columns show decomposition results for the raw achievement gaps of foreign origin (i.e., where at least one parent was born abroad) relative to native students presented in Figure 2 and Tables A1 to A3 for different starting cohorts (SC2–4) and grades (from "K1", the penultimate year of kindergarten for SC2 to "G12", i.e., grade 12, for SC4). The unexplained part measures the size of the conditional relative to the raw gap. The explained part is further decomposed into the share that is explained by kindergarten or school fixed effects ("Institutional characteristics") versus individual and family characteristics as described in Section 3.3. – *Source*: Own calculations based on NEPS data, Starting Cohorts 2–4, 2010–2018.

emerge from different reasons. Although foreign origin children do catch up to a certain extent during the early years of school, it may be the case that initial migration-induced disadvantages such as speaking a language other than the host country's at home, or being less familiar with the educational system, still have effects at later stages of the school career. Also, foreign origin students might experience some form of discrimination, e.g., by receiving less support by the teachers, or that they develop or are confronted with social norms that expect them to achieve less, all of which may result in lower performance even in standardized tests. While the conditional gaps remain substantial in most domains and stages of the school career, the gaps mostly decrease rather than increase. This demonstrates that schools are partly able to balance different backgrounds and give foreign origin students the opportunity to develop their skills in different domains over the school career, such that initial disadvantages do not add up over time.

4.2 Heterogeneity of Results and Additional Achievement Measures

The result that the specific school attended becomes as important or even more important than the role of individual and family characteristics in explaining the achievement gap between foreign origin and native students raises the question what types of schools foster the relative achievement of foreign origin students most. Therefore, Figure A3 shows raw and conditional achievement gaps for the most important types of secondary school, i.e., academic track vs. non-academic track schools for the teen cohort (SC3),[19] where 44 percent of the foreign origin students and 55 percent of natives in the balanced panel attend a school that only offers the academic track ("Gymnasium"). The results reveal that the raw gaps are smaller in the academic compared to the non-academic track schools for all different domains, which points to a selection of very high-performing foreign origin students to academic track schools. The conditional gaps are in the same range in the academic and the non-academic track, but their decrease is more pronounced in the non-academic track. Given a certain sorting of foreign origin students to academic track schools, the role of school fixed effects and socio-economic background in explaining remaining achievement gaps is then relatively limited for these high-performing foreign origin students.

To assess how the importance of track choice relates to school or neighborhood quality, I

[19]For the adolescent cohort (SC4), most students in the balanced panel that includes upper secondary education attend an academic track school. Therefore, splitting the sample by school type does not make sense for the oldest cohort.

further show in Figure A4 achievement gaps from Figure 2 where I stepwise go from raw to conditional gaps by first adding track fixed effect and then school fixed effects. The results reveal that both track choice and within-track quality account for about the same size of the difference between raw and conditional gaps, meaning that it is both important to facilitate foreign origin children to attend the academic track as well as to foster school quality in general.

Moreover, the role of the socio-economic background might be quite different for families where only one parent is foreign-born, as one native parent may be equally equipped to guide their child through school and to ensure the absence of language-related disadvantages as are two native parents. In the broader definition used above, in accordance with the literature, children with one native and one foreign-born parent are usually considered as having a foreign origin (or being 1^{st}- or 2^{nd}-generation immigrants), because disadvantages may also emerge from having a foreign-sounding name or a foreign appearance. To assess how my main results depend on the definition of immigrant children, I show results for the alternative definition of migrants (children whose parents were both born abroad) in Figure A5. The raw gaps here are up to 5 percentage points larger compared to Figure 2, which demonstrates that children with one native-born parent have much better test score results than those with two foreign-born parents. The difference of the conditional gaps compared to the main results is not as large, however, which suggests that conditional on the parents' place of birth, other family characteristics lose importance. The overall pattern of the evolution of the gaps is similar to the one observed for the main definition of migrants. Only for math test results during primary school, foreign origin children whose parents were both born abroad slightly lose ground over the years (see Panel C of Figure A5).

While the relative performance in standardized tests is a means to investigate the relative performance of foreign origin vs. native students on an objective scale, the grades in school reports eventually matter for students' obtained degrees and future career prospects. I therefore also calculate achievement gaps for these more subjective and context-dependent measures of achievement. The gaps for school report grades in German and math are shown in Figure A6 and Tables A4 to A5. The general picture is similar here to the one observed for test scores in reading or math. For the child cohort (SC2), gaps are smaller than for the older cohorts (and conditional gaps are partly slightly positive). For students from the teen cohort (SC3), a

certain degree of convergence in achievement throughout lower secondary school can be observed in both school subjects. However, grades slightly widen again in the penultimate (grade 9 for German) or ultimate year (grade 10 for math). For students from the adolescent cohort (SC4), who transition from lower to upper secondary school, a decrease in gaps for German grades can be observed. For math, the pattern shows a slight divergence. Standard errors for German and math grades are in general much larger than those for test scores as the variation in grades is limited. The scope of the gaps for grades in German and math is generally smaller than for test scores, which may be due to the fact that grades are awarded relative to peer performance, and that foreign origin students are over represented in lower tracks.

5 Conclusion

The integration of children of immigrants in the educational system is an important challenge for host countries, and is becoming more important given rising shares of students with a foreign origin. Giving each child the opportunity to unfold their potential, learn skills and eventually obtain school degrees to the same extent as their native peers is an important right for immigrants and their descendants in countries that commit to granting equal opportunities for all parts of society. The successful integration of foreign origin children in the educational system is a crucial prerequisite for later labor market success and thus decisive to improve welfare for both 1^{st}- and 2^{nd}-generation immigrants as well as host countries' native populations.

Previous literature has shown that children of immigrants still lag behind their native peers in different measures of school achievement, even if they immigrated as young children or were born in the host country. For Germany, migrant-native achievement gaps are particularly high. While most of the literature relies on cross-sectional data, I use data from a national longitudinal education study to investigate how the achievement gaps develop over the school career.

The size of the gaps is substantial, and broadly in line with results from previous cross-sectional studies. The raw gaps amount to 10 percent and beyond for most of the domains and stages of the school career. Controlling for school fixed effects and individual as well as family characteristics, i.e., measures of the socio-economic background, explains a substantial part, one to two thirds, of the gap. While both raw and conditional gaps slightly decrease over the school career (except for primary school), the importance of school and family characteristics

generally decreases as children grow older. Accordingly, the unexplained part of the decomposed achievement gap is larger towards the end of secondary schooling. Track choice is as important as further quality measures captured by school fixed effects.

The results reveal that the educational system in Germany is able to balance foreign origin students' initial disadvantages to some extent as they progress through school, which is in line with compensatory rather than accumulating effects. However, the fact that conditional gaps remain substantial in most domains and decrease less than raw gaps demonstrates that there is still scope to better support foreign origin students during their school career.

References

Alieva, A., Hildebrand, V. A., and Van Kerm, P. (2018). How does the achievement gap between immigrant and native-born pupils progress from primary to secondary education? *Luxembourg Institute of Socio-Economic Research (LISER) Working Paper Series*, 20.

Ammermueller, A. (2007). Poor background or low returns? Why immigrant students in Germany perform so poorly in the programme for international student assessment. *Education Economics*, 15(2):215–230.

Aßmann, C., Steinhauer, H. W., Kiesl, H., Koch, S., Schönberger, B., Müller-Kuller, A., Rohwer, G., Rässler, S., and Blossfeld, H.-P. (2011). 4 Sampling designs of the National Educational Panel Study: challenges and solutions. *Zeitschrift für Erziehungswissenschaft*, 14(2):51–65.

Autorengruppe Bildungsberichterstattung (2018). Bildung in Deutschland 2018. Ein indikatorengestützter Bericht mit einer Analyse zu Wirkungen und Erträgen von Bildung.

Autorengruppe Bildungsberichterstattung (2020). Bildung in Deutschland 2020. Ein indikatorengestützter Bericht mit einer Analyse zu Bildung in einer digitalisierten Welt.

Bartolucci, C. (2014). Understanding the native-immigrant wage gap using matched employer-employee data: Evidence from Germany. *Industrial and Labor Relations Review*, 67(4):1166–1202.

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

Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3):789–865.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, 8(4):436–455.

Blossfeld, H.-P., Roßbach, H.-G., and von Maurice, J. E. (2011). Education as a lifelong process. The German National Educational Panel Study (NEPS). *Zeitschrift für Erziehungswissenschaft*, Special Issue 14, Suppl 2.

Bond, T. N. and Lang, K. (2018). The black-white education scaled test-score gap in Grades K–7. *Journal of Human Resources*, 53(4):891–917.

Borjas, G. J. (1985). Assimilation, changes in cohort quality, and the earnings of immigrants. *Journal of Labor Economics*, 3(4):463–489.

Busse, A. and Gathmann, C. (2020). Free daycare policies, family choices and child development. *Journal of Economic Behavior & Organization*, 179:240–260.

Cornelissen, T., Dustmann, C., Raute, A., and Schönberg, U. (2018). Who benefits from universal child care? Estimating marginal returns to early child care attendance. *Journal of Political Economy*, 126(6):2356–2409.

Dunn, L. M. and Dunn, L. M. (1981). *Peabody picture vocabulary test-revised*. Circle Pines, MN: American Guidance Service/Pearson Education.

Dustmann, C., Frattini, T., and Lanzara, G. (2012). Educational achievement of second-generation immigrants: an international comparison. *Economic Policy*, 27(69):143–185.

Dustmann, C., Machin, S., and Schönberg, U. (2010). Ethnicity and educational achievement in compulsory schooling. *Economic Journal*, 120(546):F272–F297.

Entorf, H. and Minoiu, N. (2005). What a difference immigration policy makes: A comparison of PISA scores in Europe and traditional countries of immigration. *German Economic Review*, 6(3):355–376.

Federal Statistical Office (2020). *Bevölkerung und Erwerbstätigkeit. Bevölkerung mit Migrationshintergrund. Ergebnisse des Mikrozensus 2019 (Fachserie 1 Reihe 2.2)*. Statistisches Bundesamt, Wiesbaden.

Fischer, L., Rohm, T., Gnambs, T., and Carstensen, C. H. (2016). Linking the data of the competence tests. NEPS Survey Paper No. 1, Leibniz Institute for Educational Trajectories, National Educational Panel Study, Bamberg.

Fleurbaey, M. and Peragine, V. (2013). Ex ante versus ex post equality of opportunity. *Economica*, 80(317):118–130.

Fortin, N., Lemieux, T., and Firpo, S. (2011). Decomposition methods in economics. In *Handbook of Labor Economics*, volume 4, pages 1–102. Elsevier.

Fryer Jr., R. G. and Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. *Review of Economics and Statistics*, 86(2):447–464.

Gehrer, K., Zimmermann, S., Artelt, C., and Weinert, S. (2013). NEPS framework for assessing reading competence and results from an adult pilot study. *Journal for Educational Research Online*, 5(2):50–79.

Giannelli, G. C. and Rapallini, C. (2016). Immigrant student performance in math: Does it matter where you come from? *Economics of Education Review*, 52:291–304.

Goldin, C., Kerr, S. P., Olivetti, C., and Barth, E. (2017). The expanding gender earnings gap: Evidence from the LEHD-2000 Census. *American Economic Review: Papers and Proceedings*, 107(5):110–114.

Hanushek, E. A., Kain, J. F., and Rivkin, S. G. (2009). New evidence about Brown v. Board of Education: The complex effects of school racial composition on achievement. *Journal of Labor Economics*, 27(3):349–383.

Isphording, I. E., Piopiunik, M., and Rodríguez-Planas, N. (2016). Speaking in numbers: The effect of reading performance on math performance among immigrants. *Economics Letters*, 139:52–56.

Kerm, P. V., Yu, S., and Choe, C. (2016). Decomposing quantile wage gaps: a conditional likelihood approach. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 65(4):507–527.

Kiss, D. (2013). Are immigrants and girls graded worse? Results of a matching approach. *Education Economics*, 21(5):447–463.

Kitagawa, E. M. (1955). Components of a difference between two rates. *Journal of the American Statistical Association*, 50(272):1168–1194.

Krause, A., Rinne, U., and Schüller, S. (2015). Kick it like Özil? Decomposing the native-migrant education gap. *International Migration Review*, 49(3):757–789.

Kristen, C., Edele, A., Kalter, F., Kogan, I., Schulz, B., Stanat, P., and Will, G. (2011). 8 The education of migrants and their children across the life course. *Zeitschrift für Erziehungswissenschaft*, 14(2):121–137.

Lubotsky, D. and Kaestner, R. (2016). Do 'skills beget skills'? Evidence on the effect of kindergarten entrance age on the evolution of cognitive and non-cognitive skill gaps in childhood. *Economics of Education Review*, 53:194–206.

Merton, R. K. (1968). The Matthew Effect in science. *Science*, 159(3810):56–63.

Neumann, I., Duchhardt, C., Grüßing, M., Heinze, A., Knopp, E., and Ehmke, T. (2013). Modeling and assessing mathematical competence over the lifespan. *Journal for Educational Research Online*, 5(2):80–109.

Nguyen, H. T., Connelly, L. B., Le, H. T., Mitrou, F., Taylor, C. L., Zubrick, S. R., et al. (2020). Ethnicity differentials in academic achievements: The role of time investments. *Journal of Population Economics*, 33:1381–1418.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3):693–709.

OECD (2013). *PISA 2012 Results: Excellence through Equity: Giving Every Student the Chance to Succeed (Volume II)*. PISA. OECD: Paris.

OECD (2016). *PISA 2015 Results (Volume I): Excellence and Equity in Education*. PISA. OECD: Paris.

OECD (2018). The resilience of students with an immigrant background. Factors that shape well-being. In *OECD Reviews of Migrant Education*. OECD: Paris.

OECD (2019). *PISA 2018 Results (Volume II). Where All Students Can Succeed*. PISA, OECD: Paris.

OECD (2020). *International Migration Outlook 2020*. OECD: Paris.

Olczyk, M., Seuring, J., Will, G., and Zinn, S. (2016). Migranten und ihre Nachkommen im deutschen Bildungssystem: Ein aktueller Überblick. In *Ethnische Ungleichheiten im Bildungsverlauf*, pages 33–70. Springer.

Pohl, S. and Carstensen, C. (2012). NEPS technical report – Scaling the data of the competence tests. NEPS Working Paper No. 14, Leibniz Institute for Educational Trajectories, National Educational Panel Study, Bamberg.

Pohl, S. and Carstensen, C. H. (2013). Scaling of competence tests in the National Educational Panel Study – Many questions, some answers, and further challenges. *Journal for Educational Research Online*, 5(2):189–216.

Riphahn, R. T. (2003). Cohort effects in the educational attainment of second generation immigrants in Germany: An analysis of census data. *Journal of Population Economics*, 16(4):711–737.

Roemer, J. E. (1998). *Equality of Opportunity*. Harvard University Press, Cambridge, Mass.

Ruhose, J. and Schwerdt, G. (2016). Does early educational tracking increase migrant-native achievement gaps? Differences-in-differences evidence across countries. *Economics of Education Review*, 52:134–154.

Schnepf, S. V. (2007). Immigrants' educational disadvantage: an examination across ten countries and three surveys. *Journal of Population Economics*, 20(3):527–545.

Soto-Calvo, E. and Sánchez-Barrioluengo, M. (2016). The influence of early literacy competences on later mathematical attainment. evidence from TIMMS & PIRLS. EUR 28010 EN, Joint Research Centre (JRC) of the European Commission.

Stanat, P. and Christensen, G. (2006). *Where Immigrant Students Succeed: A Comparative Review of Performance and Engagement in PISA 2003*. OECD: Paris.

Steinhauer, H. W. and Zinn, S. (2016a). NEPS technical report for weighting: Weighting the sample of starting cohort 3 of the National Educational Panel Study (waves 1 to 3). NEPS Working Paper No. 63, Leibniz Institute for Educational Trajectories, Bamberg.

Steinhauer, H. W. and Zinn, S. (2016b). NEPS technical report for weighting: Weighting the sample of starting cohort 4 of the National Educational Panel Study (wave 1 to 6). NEPS Survey Paper No. 2, Leibniz Institute for Educational Trajectories, Bamberg.

Steinhauer, H. W., Zinn, S., Gaasch, C., and Goßmann, S. (2016). NEPS technical report for weighting: Weighting the sample of kindergarten children and Grade 1 students of the National Educational Panel Study (wave 1 to 3). NEPS Working Paper No. 66, Leibniz Institute for Educational Trajectories, Bamberg.

SVR (2013). Segregation an deutschen Schulen: Ausmaß, Folgen und Handlungsempfehlungen für bessere Bildungschancen. Sachverständigenrat deutscher Stiftungen für Integration und Migration, Berlin.

SVR (2016). Doppelt benachteiligt? Kinder und Jugendliche mit Migrationshintergrund im deutschen Bildungssystem. Sachverständigenrat deutscher Stiftungen für Integration und Migration, Berlin.

The Heckman Equation project (2021). Heckman. The economics of human potential. `https://heckmanequation.org/`. Accessed: 12/22/2020.

Weinert, S., Artelt, C., Prenzel, M., Senkbeil, M., Ehmke, T., and Carstensen, C. H. (2011). 4 Development of competencies across the life span. *Zeitschrift für Erziehungswissenschaft*, 14(2):67–86.

A Appendix

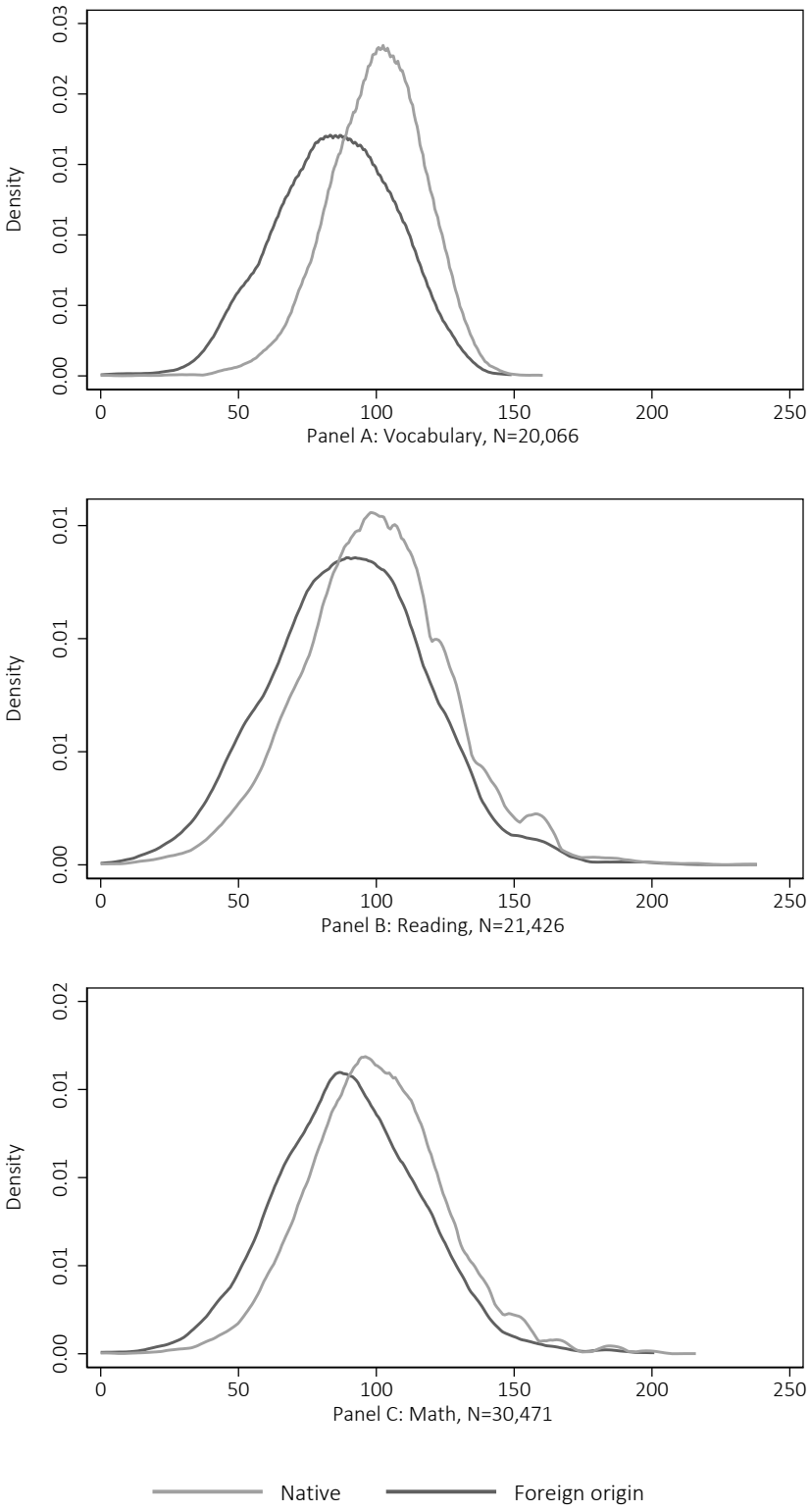

Figure A1: Distribution of Test Scores for Native vs. Foreign Origin Students

Source: Own calculations based on NEPS data, Starting Cohorts 2–4, 2010–2018.

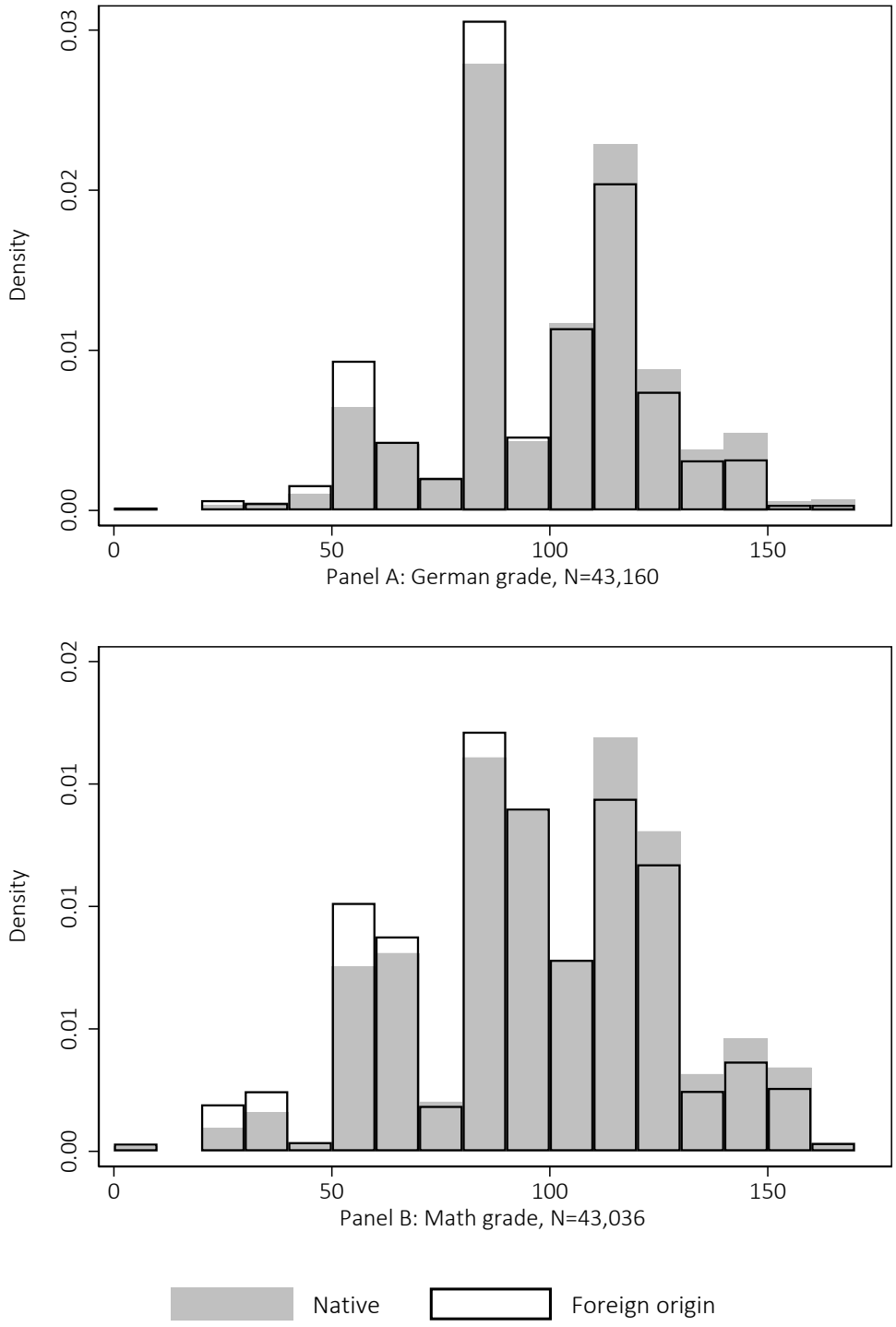

Figure A2: Distribution of School Report Grades for Native vs. Foreign Origin Students

Source: Own calculations based on NEPS data, Starting Cohorts 2–4, 2010–2018.

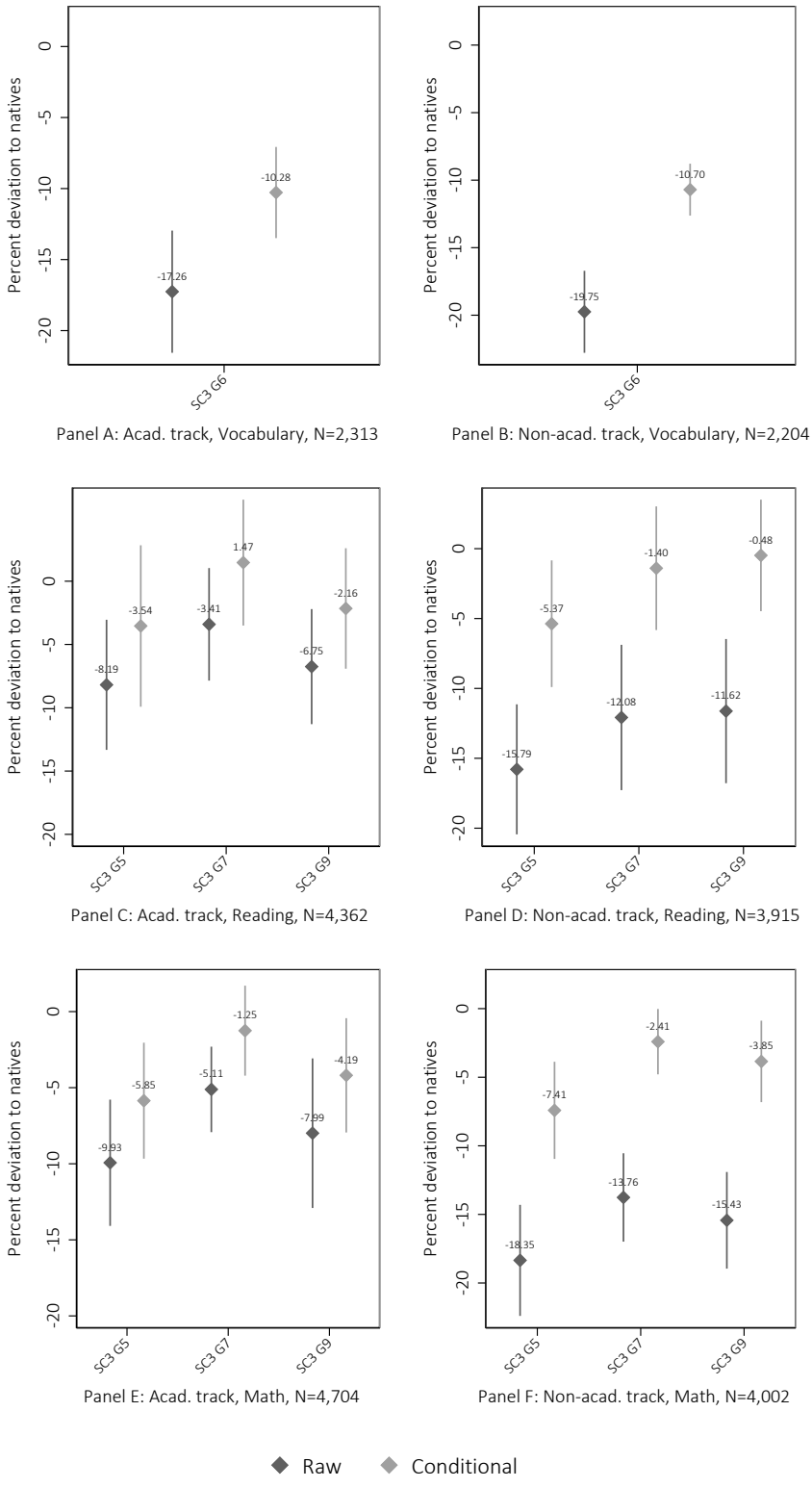

Figure A3: Migrant-Native Achievement Gaps for Test Scores: Academic vs. Non-Academic Track (for Teen Cohort)

Notes: Coefficient estimates for the academic track (left side) and non-academic tracks (right side) for raw and conditional achievement gaps of foreign origin (i.e., where at least one parent was born abroad) relative to native students for the teen cohort (SC3) for grade 5 ("G5", i.e., the first year of lower secondary school) to grade 9 ("G9"). Conditional gaps are estimated by controlling for school fixed effects and individual and family characteristics as described in Equation 1. Table 2 includes a full description of control variables. Vertical bars depict 95% CIs. – *Source*: Own calculations based on NEPS data, Starting Cohorts 2–4, 2010–2018.

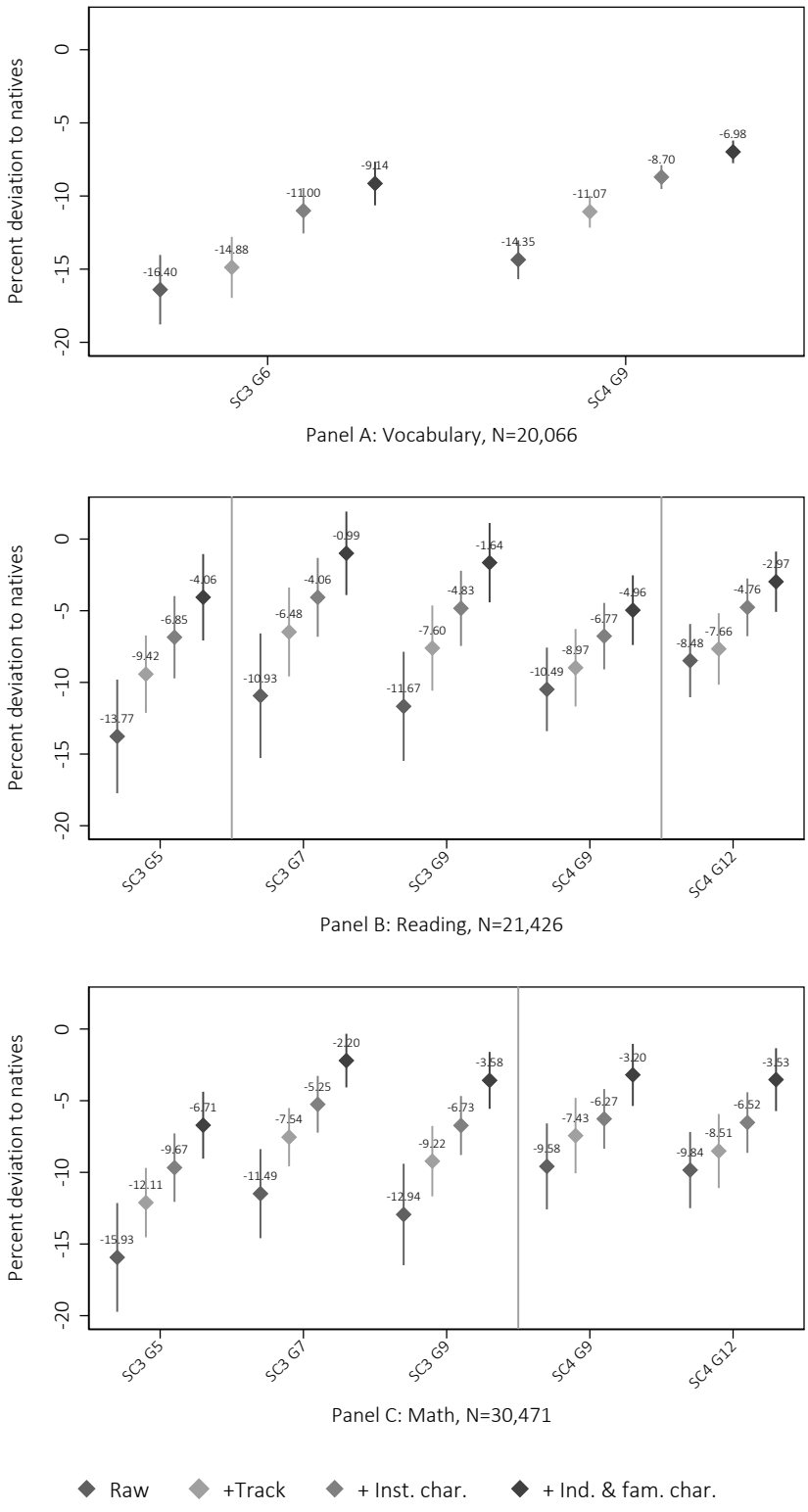

Figure A4: Migrant-Native Achievement Gaps for Test Scores: Track vs. School Fixed Effects

Notes: Coefficient estimates for raw and conditional achievement gaps of foreign origin (i.e., where at least one parent was born abroad) relative to native students for different starting cohorts (SC3–4) and grades (from "G5", i.e., grade 5, the first year of lower secondary school for SC3 to "G12", i.e., grade 12, for SC4). Raw and conditional gaps correspond to those reported in Figure 2 for SC3 and SC4, but conditional gaps here stepwise control for school track, school fixed effects ("Institutional characteristics") and individual and family characteristics as described in Equation 1. Table 2 includes a full description of control variables. Vertical bars depict 95% CIs. – *Source*: Own calculations based on NEPS data, Starting Cohorts 3–4, 2010–2018.

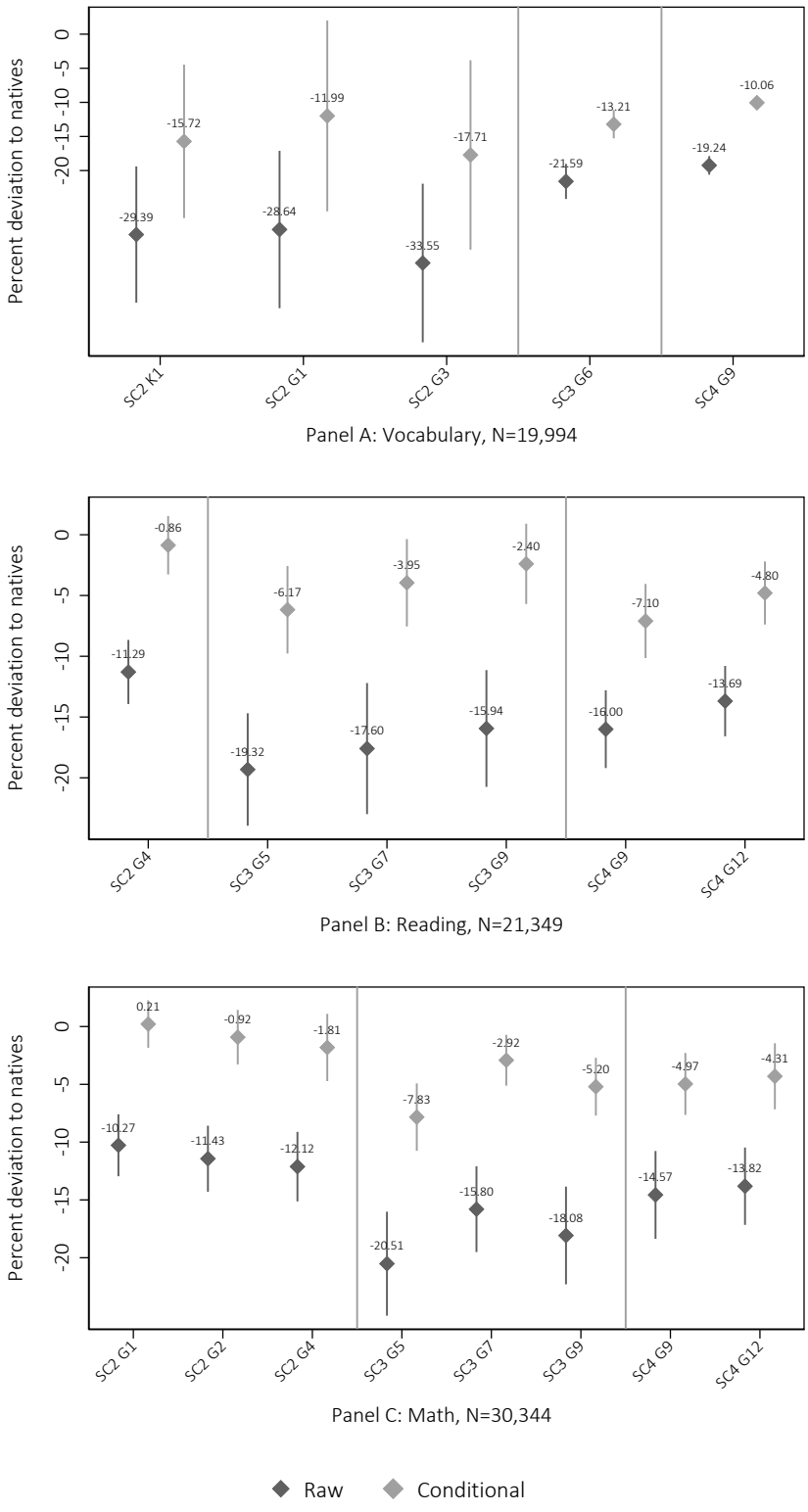

Figure A5: Migrant-Native Achievement Gaps for Test Scores: Alternative Definition of Migrants

Notes: Coefficient estimates for raw and conditional achievement gaps of foreign origin relative to native students for different starting cohorts (SC2–4) and grades (from "G1", the first year of primary school for SC2 to "G12", i.e., grade 12, for SC4). Conditional gaps are estimated by controlling for kindergarten or school fixed effects and individual and family characteristics as described in Equation 1. Table 2 includes a full description of control variables. Vertical bars depict 95% CIs. In contrast to Figure 2 where immigrant children from families with one foreign-born and one native parent were considered as having a foreign origin, these children here are treated as natives. – *Source*: Own calculations based on NEPS data, Starting Cohorts 2–4, 2010–2018.

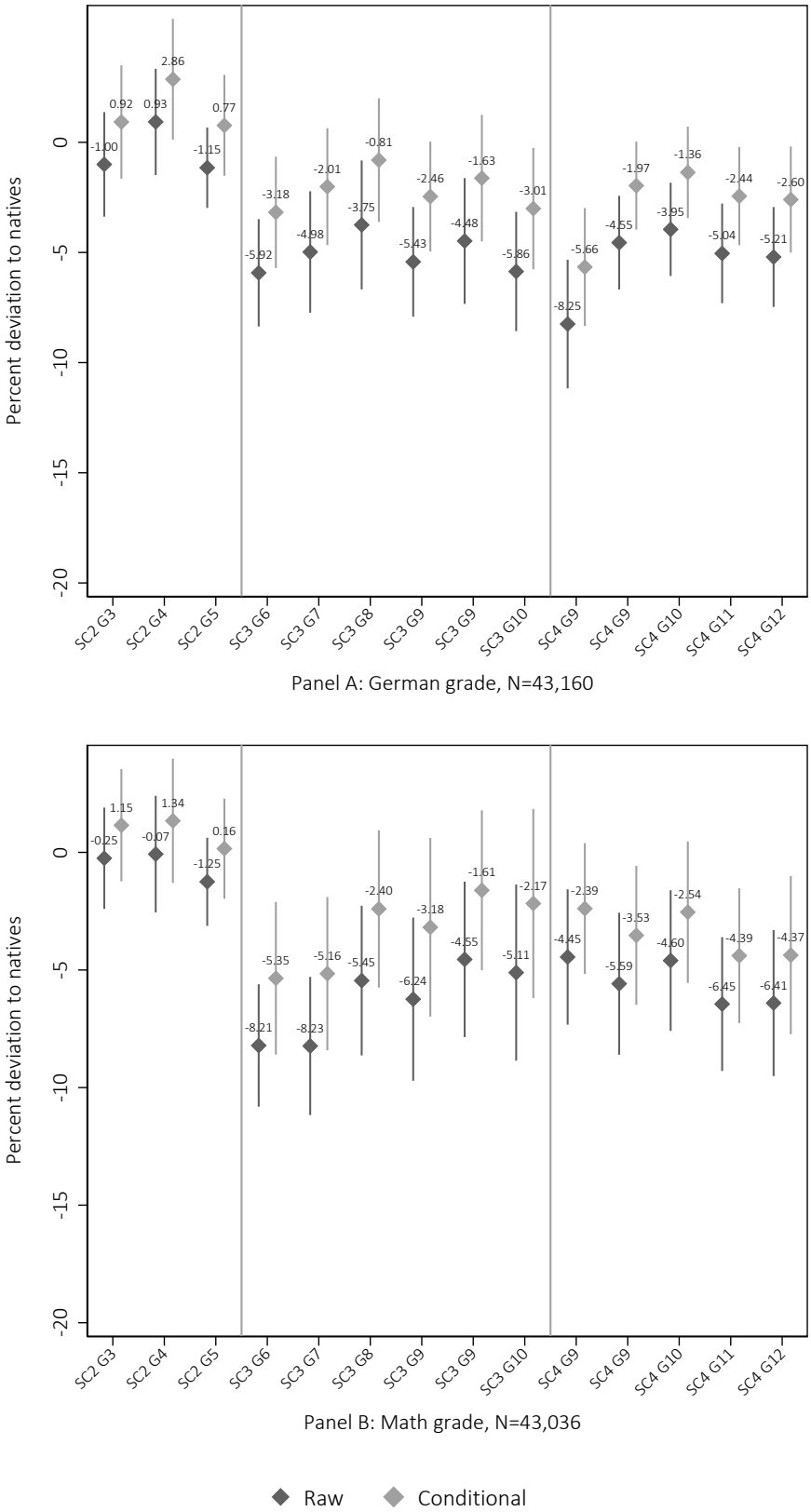

Figure A6: Migrant-Native Achievement Gaps for School Report Grades

Notes: Coefficient estimates for raw and conditional achievement gaps of foreign origin (i.e., where at least one parent was born abroad) relative to native students for different starting cohorts (SC2–4) and grades (from "G3", the third year of primary school for SC2 to "G9", i.e., grade 9, for SC4). Conditional gaps are estimated by controlling for school fixed effects and individual and family characteristics as described in Equation 1 and presented in Tables A4 to A5. Table 2 includes a full description of control variables. Vertical bars depict 95% CIs. – *Source*: Own calculations based on NEPS data, Starting Cohorts 2–4, 2010–2018.

Table A1: Migrant-Native Achievement Gaps in Vocabulary

	Raw Coeff./SE	+School FE Coeff./SE	+Ind. & Fam. Char. Coeff./SE
SC2 K1	−25.32***	−24.42***	−12.46***
	(4.16)	(4.89)	(4.43)
SC2 G1	−21.69***	−16.73***	−5.81
	(5.01)	(5.93)	(5.71)
SC2 G3	−23.67***	−18.71***	−8.19
	(5.32)	(6.34)	(6.00)
SC3 G6	−16.40***	−11.00***	−9.14***
	(1.21)	(0.79)	(0.76)
SC4 G9	−14.35***	−8.70***	−6.98***
	(0.68)	(0.42)	(0.39)
Adjusted R^2	0.11	0.39	0.44
Institutions	813	813	813
Observations	20,066	20,066	20,066

Notes: The table shows gaps in Vocabulary test scores of foreign origin (i.e., where at least one parent was born abroad) relative to native students for different starting cohorts and grades (from "K1", the penultimate year of kindergarten for SC2 to "G9", i.e., grade 9, for SC4). The dependent variable is scaled to having a mean of 100 for natives and a minimum of 0 in each wave. Column (1) includes raw achievement gaps without any control variables. The following columns step wise add controls for (2) kindergarten or school fixed effects, and (3) individual and family characteristics: a dummy for girls, 5 categories for the number of books at home, and 4–5 categories for each of the mother's and father's education level and occupation status (see Table 2 for a full description of control variables). – Asterisks indicate p-values according to: *** p<0.01, ** p<0.05, * p<0.1. – *Source*: Own calculations based on NEPS data, Starting Cohorts 2–4, 2010–2018.

Table A2: Migrant-Native Achievement Gaps in Reading

	Raw Coeff./SE	+School FE Coeff./SE	+Ind. & Fam. Char. Coeff./SE
SC2 G4	-5.37^{***}	-4.43^{***}	1.36
	(1.06)	(0.98)	(0.96)
SC3 G5	-13.77^{***}	-6.85^{***}	-4.06^{***}
	(2.02)	(1.46)	(1.54)
SC3 G7	-10.93^{***}	-4.06^{***}	-0.99
	(2.21)	(1.40)	(1.48)
SC3 G9	-11.67^{***}	-4.83^{***}	-1.64
	(1.94)	(1.34)	(1.41)
SC4 G9	-10.49^{***}	-6.77^{***}	-4.96^{***}
	(1.49)	(1.18)	(1.24)
SC4 G12	-8.48^{***}	-4.76^{***}	-2.97^{***}
	(1.30)	(1.03)	(1.07)
Adjusted R^2	0.02	0.22	0.25
Institutions	694	694	694
Observations	21,426	21,426	21,426

Notes: The table shows gaps in Reading comprehension test scores of foreign origin (i.e., where at least one parent was born abroad) relative to native students for different starting cohorts and grades (from "G4", i.e., grade 4, the last year of primary school for SC2 to "G12", i.e., grade 12, the (pen-)ultimate year of upper secondary schooling, for SC4). The dependent variable is scaled to having a mean of 100 for natives and a minimum of 0 in each wave. Column (1) includes raw achievement gaps without any control variables. The following columns step wise add controls for (2) school fixed effects, and (3) individual and family characteristics: a dummy for girls, 5 categories for the number of books at home, and 4–5 categories for each of the mother's and father's education level and occupation status (see Table 2 for a full description of control variables). – Asterisks indicate p-values according to: *** p<0.01, ** p<0.05, * p<0.1. – *Source*: Own calculations based on NEPS data, Starting Cohorts 2–4, 2010–2018.

Table A3: Migrant-Native Achievement Gaps in Math

	Raw Coeff./SE	+School FE Coeff./SE	+Ind. & Fam. Char. Coeff./SE
SC2 G1	-7.50^{***}	-6.22^{***}	-0.81
	(0.97)	(0.86)	(0.87)
SC2 G2	-7.88^{***}	-6.60^{***}	-1.18
	(1.05)	(0.94)	(0.93)
SC2 G4	-7.12^{***}	-5.84^{***}	-0.53
	(1.18)	(1.15)	(1.10)
SC3 G5	-15.93^{***}	-9.67^{***}	-6.71^{***}
	(1.93)	(1.22)	(1.19)
SC3 G7	-11.49^{***}	-5.25^{***}	-2.20^{**}
	(1.58)	(1.01)	(0.95)
SC3 G9	-12.94^{***}	-6.73^{***}	-3.58^{***}
	(1.80)	(1.05)	(1.01)
SC4 G9	-9.58^{***}	-6.27^{***}	-3.20^{***}
	(1.53)	(1.06)	(1.10)
SC4 G12	-9.84^{***}	-6.52^{***}	-3.53^{***}
	(1.35)	(1.08)	(1.12)
Adjusted R^2	0.03	0.23	0.30
Institutions	674	674	674
Observations	30,471	30,471	30,471

Notes: The table shows gaps in Math test scores of foreign origin (i.e., where at least one parent was born abroad) relative to native students for different starting cohorts and grades (from "G1", i.e., grade 1, the first year of primary school for SC2 to "G12", i.e., grade 12, the (pen-)ultimate year of upper secondary schooling, for SC4). The dependent variable is scaled to having a mean of 100 for natives and a minimum of 0 in each wave. Column (1) includes raw achievement gaps without any control variables. The following columns step wise add controls for (2) school fixed effects, and (3) individual and family characteristics: a dummy for girls, 5 categories for the number of books at home, and 4–5 categories for each of the mother's and father's education level and occupation status (see Table 2 for a full description of control variables). – Asterisks indicate p-values according to: *** p<0.01, ** p<0.05, * p<0.1. – *Source*: Own calculations based on NEPS data, Starting Cohorts 2–4, 2010–2018.

Table A4: Migrant-Native Achievement Gaps in German Grade

	Raw Coeff./SE	+School FE Coeff./SE	+Ind. & Fam. Char. Coeff./SE
SC2 G3	−1.00	−1.97	0.92
	(1.21)	(1.25)	(1.31)
SC2 G4	0.93	−0.04	2.86**
	(1.23)	(1.31)	(1.40)
SC2 G5	−1.15	−2.12**	0.77
	(0.93)	(1.05)	(1.17)
SC3 G6	−5.92***	−4.36***	−3.18**
	(1.24)	(1.24)	(1.29)
SC3 G7	−4.98***	−3.41***	−2.01
	(1.40)	(1.32)	(1.35)
SC3 G8	−3.75**	−2.18	−0.81
	(1.49)	(1.45)	(1.43)
SC3 G9	−5.43***	−3.90***	−2.46*
	(1.27)	(1.26)	(1.27)
SC3 G9	−4.48***	−2.96**	−1.63
	(1.45)	(1.47)	(1.46)
SC3 G10	−5.86***	−4.34***	−3.01**
	(1.38)	(1.45)	(1.40)
SC4 G9	−8.25***	−8.19***	−5.66***
	(1.49)	(1.36)	(1.36)
SC4 G9	−4.55***	−4.50***	−1.97*
	(1.08)	(1.00)	(1.02)
SC4 G10	−3.95***	−3.89***	−1.36
	(1.08)	(1.03)	(1.06)
SC4 G11	−5.04***	−4.99***	−2.44**
	(1.15)	(1.09)	(1.14)
SC4 G12	−5.21***	−5.14***	−2.60**
	(1.15)	(1.23)	(1.23)
Adjusted R^2	0.01	0.10	0.18
Institutions	728	728	728
Observations	43,160	43,160	43,160

Notes: The table shows gaps in German grades of foreign origin (i.e., where at least one parent was born abroad) relative to native students for different starting cohorts and grades (from "G3", i.e., grade 3, the penultimate year of primary school for SC2 to "G12", i.e., grade 12, the (pen-)ultimate year of upper secondary schooling, for SC4). The dependent variable is scaled to having a mean of 100 for natives and a minimum of 0 in each wave. Column (1) includes raw achievement gaps without any control variables. The following columns step wise add controls for (2) school fixed effects, and (3) individual and family characteristics: a dummy for girls, 5 categories for the number of books at home, and 4–5 categories for each of the mother's and father's education level and occupation status (see Table 2 for a full description of control variables). – Asterisks indicate p-values according to: *** p<0.01, ** p<0.05, * p<0.1. – *Source*: Own calculations based on NEPS data, Starting Cohorts 2–4, 2010–2018.

Table A5: Migrant-Native Achievement Gaps in Math Grade

	Raw Coeff./SE	+School FE Coeff./SE	+Ind. & Fam. Char. Coeff./SE
SC2 G3	−0.25	−1.53	1.15
	(1.10)	(1.18)	(1.22)
SC2 G4	−0.07	−1.35	1.34
	(1.26)	(1.31)	(1.34)
SC2 G5	−1.25	−2.53**	0.16
	(0.95)	(1.02)	(1.08)
SC3 G6	−8.21***	−8.16***	−5.35***
	(1.33)	(1.45)	(1.65)
SC3 G7	−8.23***	−8.18***	−5.16***
	(1.50)	(1.59)	(1.66)
SC3 G8	−5.45***	−5.40***	−2.40
	(1.62)	(1.68)	(1.70)
SC3 G9	−6.24***	−6.22***	−3.18
	(1.77)	(1.93)	(1.93)
SC3 G9	−4.55***	−4.53**	−1.61
	(1.68)	(1.76)	(1.73)
SC3 G10	−5.11***	−5.09**	−2.17
	(1.91)	(2.04)	(2.05)
SC4 G9	−4.45***	−4.44***	−2.39*
	(1.47)	(1.45)	(1.42)
SC4 G9	−5.59***	−5.57***	−3.53**
	(1.54)	(1.53)	(1.51)
SC4 G10	−4.60***	−4.59***	−2.54*
	(1.52)	(1.51)	(1.53)
SC4 G11	−6.45***	−6.44***	−4.39***
	(1.45)	(1.43)	(1.46)
SC4 G12	−6.41***	−6.42***	−4.37**
	(1.58)	(1.68)	(1.71)
Adjusted R^2	0.00	0.07	0.09
Institutions	727	727	727
Observations	43,036	43,036	43,036

Notes: The table shows gaps in Math grades of foreign origin (i.e., where at least one parent was born abroad) relative to native students for different starting cohorts and grades (from "G3", i.e., grade 3, the penultimate year of primary school for SC2 to "G12", i.e., grade 12, the (pen-)ultimate year of upper secondary schooling, for SC4). The dependent variable is scaled to having a mean of 100 for natives and a minimum of 0 in each wave. Column (1) includes raw achievement gaps without any control variables. The following columns step wise add controls for (2) school fixed effects, and (3) individual and family characteristics: a dummy for girls, 5 categories for the number of books at home, and 4–5 categories for each of the mother's and father's education level and occupation status (see Table 2 for a full description of control variables). – Asterisks indicate p-values according to: *** p<0.01, ** p<0.05, * p<0.1. – *Source*: Own calculations based on NEPS data, Starting Cohorts 2–4, 2010–2018.