How Ethnic Achievement Gaps Vary Across European Countries

Richard Nennstiel

1 Introduction

The educational attainment of students from immigrant backgrounds remains a central preoccupation for sociologists of education and a critical policy concern across Europe. A vast body of comparative research has consistently documented the persistence of ethnic penalties in educational outcomes, even when comparing students from similar socio-economic strata (Heath, 2014; Triventi et al., 2022). While this literature has provided invaluable insights into the structural barriers and integration processes affecting established minority groups, particularly the second generation (Heath et al., 2008), the European context has been fundamentally reshaped over the last decade. The period from 2014 onwards has been characterised by new, large-scale migration flows, including a significant component of forced migration (Kogan & Kalter, 2020). This development presents host-country education systems with a distinct and pressing challenge: the integration of newly arrived students, many of whom have experienced interrupted schooling and significant trauma (Glinka & Winkler, 2024; Olczyk et al., 2025; Schipolowski et al., 2021). Crucially, this population represents a highly heterogeneous stratum, ranging from intra-regional migrants to forced migrants from conflict zones. For the latter group, pre-migration educational biographies are frequently characterised by institutional disruptions and restricted access to schooling—both in countries of origin and during transit—resulting in distinct educational trajectories that significantly condition their academic integration.

Despite the profound policy relevance of this challenge, large-scale, cross-national comparative research on the educational performance of this newly arrived

R. Nennstiel (✉)
Department of Sociology of Education, University of Bern, Bern, Switzerland

Institute of Social Sciences, University of Lausanne, Lausanne, Switzerland
e-mail: richard.nennstiel@unil.ch

O. Winkler, M. Olczyk (eds.), *Educational Integration of Refugees and Newly Arrived Immigrants*, https://doi.org/10.1007/978-3-032-25346-0_8

cohort has been scarce. This scarcity is largely attributable to a lack of timely, representative, and comparative data. The 2022 wave of the Programme for International Student Assessment (PISA) (OECD, 2023a) provides a major empirical opportunity to address this lacuna. By distinguishing students based on their arrival date, the PISA 2022 data (OECD, 2023b) allow us to isolate the cohort of 15-year-olds who arrived in their respective host countries in or after 2014. This study leverages this dataset to provide a comparative portrait of the mathematical proficiency gaps between these newly arrived students and their native peers across 16 European host countries. Analysing mathematical literacy is essential, as it constitutes a key competence that functions as a decisive factor for career choices and subsequent career prospects (Gutfleisch & Nennstiel, 2025; Ritchie & Bates, 2013).

This chapter addresses three central research questions. First, what is the magnitude of the unadjusted performance gap between newly arrived immigrant students and native students, and how does this gap vary across the European landscape? Second, to what extent are these observed gaps attributable to compositional differences—specifically, disadvantages in socio-economic background and the lack of host-country-specific human capital, proxied by home language? Third, after accounting for these individual-level factors, do institutional features of the host countries' education systems, namely the degree of educational stratification, moderate the remaining performance gaps?

This chapter proceeds as follows. In the next section, we will elaborate on the theoretical framework guiding this analysis, drawing on concepts from migration studies, assimilation and integration theories, and the sociology of educational stratification to develop our hypotheses. Following this, we detail the methodological approach, describing the PISA 2022 data, our sample of 16 European host countries, and the operationalisation of our key variables. The core empirical section then presents the results in a stepwise fashion: we begin with a descriptive overview, quantify the unadjusted proficiency gaps, and then systematically decompose these gaps by controlling for compositional factors. The empirical analysis culminates in a macro-level test of our institutional hypothesis. Finally, the chapter concludes with a comprehensive discussion of these findings, their limitations, and their implications.

2 Theoretical Considerations

The observation that students with an immigrant background often exhibit lower average test scores than their native-majority peers is one of the most robust findings in the sociology of education (Heath, 2014). This initial analysis, therefore, does not seek to merely "discover" these well-documented gaps. Instead, it positions them as a baseline from which a more nuanced, three-tiered theoretical inquiry proceeds.

First, classical assimilation and integration theories posit that performance differentials should attenuate over time, both with increasing duration of residence and

across generations (Fleischmann & Kristen, 2014; Laubenthal, 2023; Nennstiel, 2022c). The mechanisms are twofold: the gradual acquisition of host-country-specific human capital, most critically language proficiency (Schipolowski et al., 2021), and the intergenerational progression of socio-economic integration (Drouhot & Nee, 2019). This framework leads to a clear set of graduated expectations: performance gaps should be largest for the most recently arrived first-generation students, smaller for first-generation students with a longer duration of residence, and smallest (though not necessarily non-existent) for the second generation, who have benefited from full exposure to the host country's educational institutions (Heath et al., 2008).

Second, any comparative analysis of these gaps must confront the profound methodological challenge of compositional effects. The observed performance of any immigrant group is inextricably linked to its selectivity profile (Schmidt et al., 2021; Spörlein et al., 2020; Spörlein & Kristen, 2019). The ethnic test score gap is, therefore, a composite measure that conflates genuine integration challenges with pre-migration disparities. Two such determinants are paramount. The first is socio-economic status, operationalised in PISA as the ESCS index (Avvisati, 2020), which remains the single most powerful predictor of academic achievement globally (Pokropek et al., 2015). If newly arrived immigrants are on average disproportionately concentrated in the lower strata of the socio-economic hierarchy, a significant portion of their performance gap may be an expression of class inequality rather than a unique migrant-specific penalty. It is crucial to acknowledge the substantial heterogeneity inherent in this newly arrived immigrant population. This group encompasses a broad spectrum ranging from highly qualified labour migrants and intra-European movers to refugees. Consequently, any observed concentration in lower socio-economic strata represents an aggregate average that masks significant internal diversity. The second is linguistic capital. A lack of proficiency in the language of instruction is a primary barrier to academic learning (Schipolowski et al., 2021), and the prevalence of speaking the test language at home serves as a powerful proxy for this dimension of human capital. It is thus theoretically essential to statistically account for these compositional factors—along with other demographic controls such as age (Urruticoechea et al., 2021) and gender (Penner, 2008)—to isolate the net gap that can be more plausibly attributed to processes occurring within the host society.

Third, after accounting for these individual-level compositional differences, we theorise that the remaining cross-national variance in performance gaps is systematically related to the institutional structures of the host-country education systems (Triventi et al., 2016). The most debated institutional feature in this regard is educational stratification, often operationalised as the degree and timing of curricular tracking (Gamoran, 1987). Highly stratified systems, which sort students into distinct academic and vocational tracks at an early age, are argued to exacerbate inequality. They may misallocate minority students by relying on assessments that are culturally or linguistically biased (Nennstiel & Gilgen, 2024), and once sorted, students in lower tracks are exposed to less rigorous curricula and negative peer effects, thereby compounding initial disparities (Nennstiel, 2022b; Terrin & Triventi,

2022). While the precise mechanisms of this effect are subject to intense scholarly debate (Heisig & Matthewes, 2022), the prevailing hypothesis suggests that highly tracked systems are less able to mitigate initial disparities. Consequently, they should impose a larger institutional penalty on newly arrived students, whose linguistic and social constrains are most acute upon entry.

Regarding the specific domain of mathematics, observed competence disparities between native and immigrant students are theoretically likely to be driven by several distinct mechanisms. For one, despite the abstract nature of the subject, mathematical acquisition remains linguistically mediated; thus, limited familiarity with the language of instruction may constrain the comprehension of complex tasks. Furthermore, performance differentials may reflect pre-migration variations in instructional quality and curricular standards within the countries of origin. Additionally, competence development is frequently stratified by socioeconomic background and subsequent track placement. Based on these considerations, we derive the following four hypotheses.

Building on the robust empirical evidence on the competence disparities between native and immigrant students we first establish a baseline expectation regarding the general performance hierarchy (Heath, 2014).

H1: Students with an immigrant background exhibit lower average test performance than native students.

Building on the temporal dimension of integration (Fleischmann & Kristen, 2014), we distinguish between groups based on their structural exposure to the host society. Native and second-generation students share the distinct advantage of being born in the destination country and thus benefit from lifelong institutional socialisation. In contrast, first-generation immigrants have necessarily resided there for only limited periods. This creates fundamental unequal starting conditions, implying that the depth of the performance disadvantage will correspond to the extent of this exposure gap.

H2: These test score gaps attenuate across generations and with duration of residence. The largest gaps are expected for newly arrived students (post-2014), while the smallest gaps are expected for the second generation.

Given the pronounced heterogeneity in socio-economic and linguistic resources (Schmidt et al., 2021; Spörlein et al., 2020; Spörlein & Kristen, 2019), the observed disadvantage is likely driven by compositional factors rather than migration status alone. We therefore posit that accounting for these unevenly distributed resources will explain a substantial portion of the raw performance differential (Pokropek et al., 2015; Schipolowski et al., 2021).

H3: The observed test score gaps between newly arrived and native students will decrease substantially after controlling for socio-economic status and language spoken at home.

Finally, we examine the macro-level impact of institutional differentiation (Gamoran, 1987; Triventi et al., 2016). In highly stratified education systems, the structural imperative to sort students into distinct curricular pathways creates specific vulnerabilities for newly arrived immigrants. Regardless of their age at arrival, these students face a significantly higher probability of being assigned to

lower-track environments, often due to initial socio-economic and linguistic barriers or diagnostic biases. This institutional segregation exposes them to less demanding curricula and differentiated instructional quality, thereby entrenching performance disparities more severely than in comprehensive systems.

H4: Among countries, those with more highly stratified education systems (as measured by the tracking index; Bol et al., 2014) exhibit larger *adjusted* test score gaps between newly arrived and native students.

3 Data and Methods

The empirical foundation for the present investigation is drawn from the 2022 wave of the PISA (OECD, 2023a, 2023b). As a seminal large-scale international survey, PISA provides robust, nationally representative data on the competencies of 15-year-old students.[1] It employs a complex, stratified sampling design methodology to ensure the rigorous comparability of educational outcomes across diverse national systems. For the purposes of our analysis, we have selectively pooled data from the participating European nations. The primary inclusion criterion for this study was the presence of a sufficient subsample of newly arrived immigrants; specifically, we retained only those countries whose datasets included a minimum threshold of 150 students who had immigrated into the respective host country subsequent to the year 2014. This threshold is methodologically necessary to ensure adequate statistical power for the planned subgroup analyses. Consequently, our comparative investigation is predicated upon the data from the following 16 countries (N = 138,782): Austria, Belgium, Switzerland, Germany, Denmark, Finland, France, Great Britain, Ireland, Malta, Norway, Portugal, Spain, Italy, the Netherlands, and Slovenia.

The operationalisation of the variables central to our analytical models proceeds as follows. Our primary dependent variable, mathematical proficiency, is operationalised using the PISA test scores. To accurately represent the latent trait of student ability, we employ all ten plausible values provided in the dataset. This standard methodological approach for large-scale assessments properly accounts for the measurement uncertainty inherent in proficiency estimation by treating the scores as a probability distribution rather than a single point estimate. The gender of the students is derived from their self-reported information within the student questionnaire and is operationalised as a dichotomous variable (boy/girl). A precise measure of student age is calculated in years, utilising the provided data on their month and year of birth. Socio-economic background is operationalised using the PISA Index of Economic, Social and Cultural Status (ESCS) (Avvisati, 2020). The ESCS is a multifaceted composite measure, meticulously constructed by the OECD, which

[1] While PISA technical standards permit participating countries to exclude students who have resided in the host country for less than 1 year, typically on the grounds of insufficient language proficiency, this exclusion is discretionary. Empirically, the dataset used in this analysis includes students across all sampled countries who reported immigrating in 2021 or 2022.

integrates information on parental education, parental occupation, and various indicators of household material and cultural capital. Within our models, higher values on this continuous index signify a higher socio-economic status and, by extension, a domestic environment richer in educationally relevant resources.

The conceptualisation of immigrant status, our key independent variable, requires precise definition. We first establish a primary dichotomy between native students (defined as students with both parents and the student themselves being born in the country of assessment) and immigrant students (defined as students where either the child or at least one parent was born abroad). Within this broad immigrant cohort, we introduce a crucial distinction based on generational status. Second-generation students are operationalised as those who were born in the host country but whose parents were both born elsewhere. First-generation students are those who were themselves born outside the host country. Given the specific theoretical focus of this study, we further disaggregate this first-generation group based on their chronology of arrival. We distinguish between those who immigrated *prior to 2014* and those who arrived *in or after 2014*. This temporal disaggregation is methodologically purposeful, as it allows us to isolate and place a specific analytical focus on newly migrated students. The selection of 2014 as a threshold is sociologically informed, intended to reflect the significant increase in forced migration flows into Europe, particularly from the Middle East, which intensified from that year onwards.

It is imperative, however, to acknowledge the clear limitations of the PISA data in this regard. The dataset does not permit the direct identification of refugee status, as students were not explicitly surveyed on their legal or humanitarian grounds for migration.[2] Furthermore, attempting to cluster immigrant students by specific countries or regions of origin presents insurmountable challenges to a comparative research design. This is because the national questionnaires vary significantly in how they categorise country of origin; which specific countries are listed and which are aggregated into a generic "other countries" category is not standardised across the participating nations. Consequently, our analytical framework is necessarily constrained to differentiating only by immigrant generation and date of arrival. We must forego a more granular analysis that accounts for refugee status or specific origin-country constellations, acknowledging this as an unavoidable data-imposed constraint. The language spoken at home is captured via a binary indicator, derived from student self-reports, distinguishing between students who predominantly speak the language of instruction (i.e., the language of the PISA test) in their domestic setting and those who predominantly speak another language. Additionally, we incorporate a further binary indicator to capture whether at least one parent was born in the host country.

[2] Despite this limitation, indirect indicators strongly suggest that refugees are included in the sample of newly arrived students. Analysis of countries providing detailed origin information confirms the presence of adolescents from nations such as Syria, Iraq, and Afghanistan, as well as those speaking languages such as Arabic. Thus, the category of newly arrived immigrants likely contains a relevant share of students with a refugee background.

To handle missing data, our analytical strategy relies on listwise deletion, often referred to as complete case analysis. This approach entails the exclusion of any student who exhibits a missing value on *any* of the model variables specified above. This methodological decision is bolstered by the low incidence of missingness; across the pooled sample, only 6.6% of students presented with at least one missing value on these key variables (see Table 1 in the Appendix for valid cases per variable and complete cases included in the analysis by country). This procedure results in a final analytical sample of N = 129,727. It is pertinent to note that the country-specific case numbers within this final sample are heterogeneous, varying substantially from a minimum of N = 2879 in Malta to a maximum of N = 29,059 in Spain.

All statistical computations were executed using Stata, version 19. To adequately and rigorously address the methodological complexities inherent in the PISA dataset, our analyses are implemented using the specific Stata package *repest* (Avvisati & Keslair, 2014). The use of this package, which was expressly designed for analysing international large-scale assessment data, is essential to account for the complex survey design. It correctly manages two fundamental data features simultaneously: first, the complex sampling structure (which involves stratification and clustering, requiring the use of replication weights for accurate variance estimation), and second, the aforementioned use of plausible values to ensure unbiased estimation of proficiency statistics and regression coefficients.

The initial phase of our empirical analysis is descriptive (see Fig. 1). We present a comparative visualisation of the compositional demography of each national sample. Specifically, Fig. 1 illustrates the relative proportions (i.e., calculated

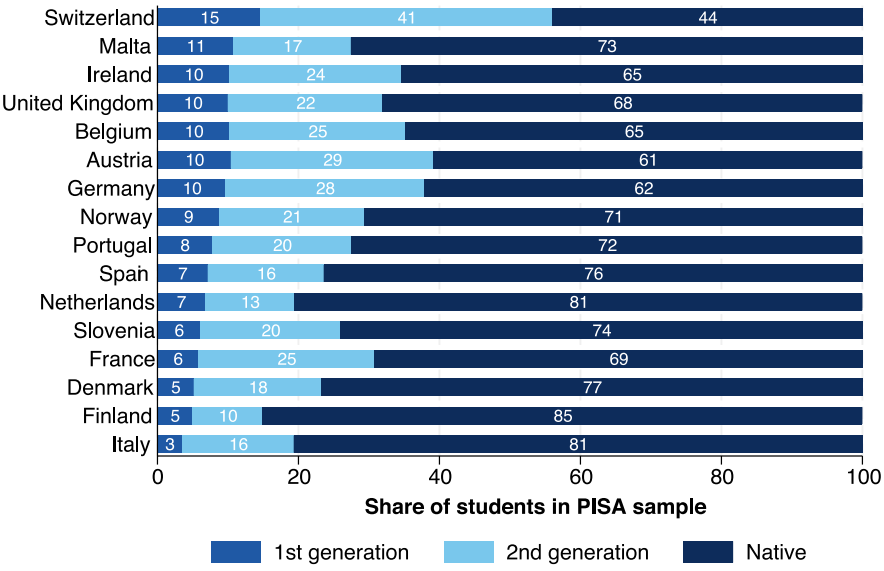

Fig. 1 Share of native, first- and second-generation immigrant students in the sample, across countries. Source: PISA 2022, weighted analysis, our own calculations

percentages) of native students and the distinct immigrant groups (first and second generation) within each of the 16 countries. This provides an essential contextual overview of the varying demographic landscapes in which educational integration is occurring.

Following this descriptive mapping, we proceed to quantify the magnitude of raw outcome disparities (see Fig. 2). This figure plots the unadjusted test score gaps in mathematical literacy between native students (the reference category) and the various immigrant groups for each nation. Crucially, in this visualisation, the first-generation cohort is disaggregated by arrival date (pre- and post-2014). These plotted gaps are derived from a series of country-specific OLS regression models in which the sole predictors are the dummy variables for immigrant group status. Therefore, the values represent the unadjusted regression coefficients, indicating the raw performance differential relative to the native reference group.

The third analytical step addresses the crucial sociological question of socio-economic selectivity among migrant populations. To investigate how newly arrived first-generation immigrants (post-2014) compare socio-economically to their native peers, we calculate and plot the mean ESCS scores for both native students and this specific newly arrived immigrant subgroup (see Fig. 3). This visualisation serves a dual analytical purpose: it facilitates a within-country comparison of the relative social positioning of these immigrant families within their host society's hierarchy, and it also enables a cross-national comparison, providing salient insights into the variation in socio-economic integration (or marginalisation) across different national contexts.

Complementing the socio-economic profile, our fourth step (see Fig. 4) examines a key indicator of cultural and linguistic integration. Here, we plot the

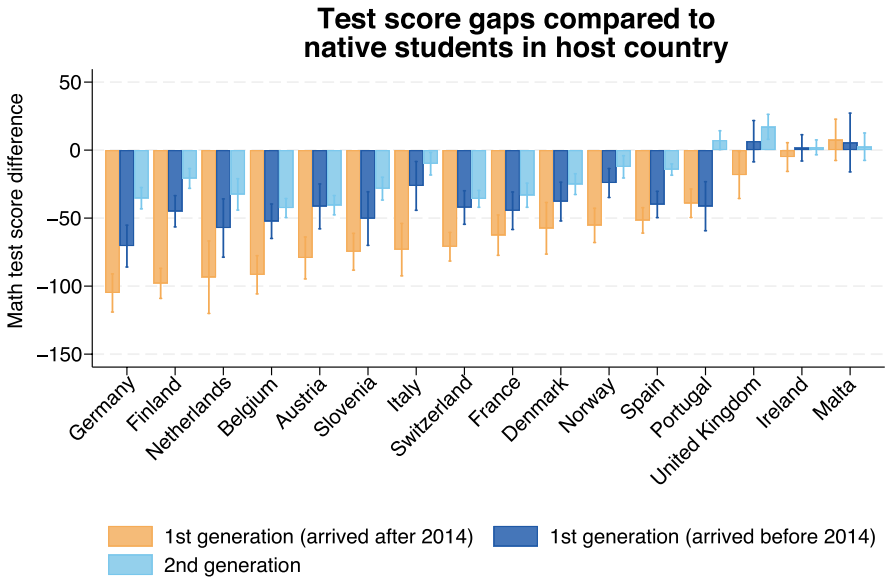

Fig. 2 Test score gaps across immigrant student groups compared to native students. Source: PISA 2022, weighted analysis, our own calculations

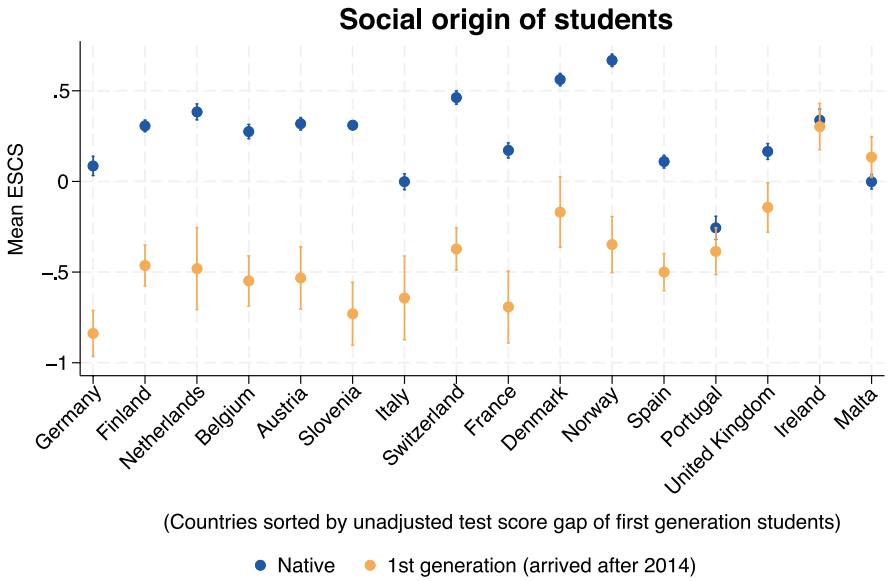

Fig. 3 Mean ESCS (social origin measure) for native students and newly arrived immigrant students, across countries. Source: PISA 2022, weighted analysis, our own calculations

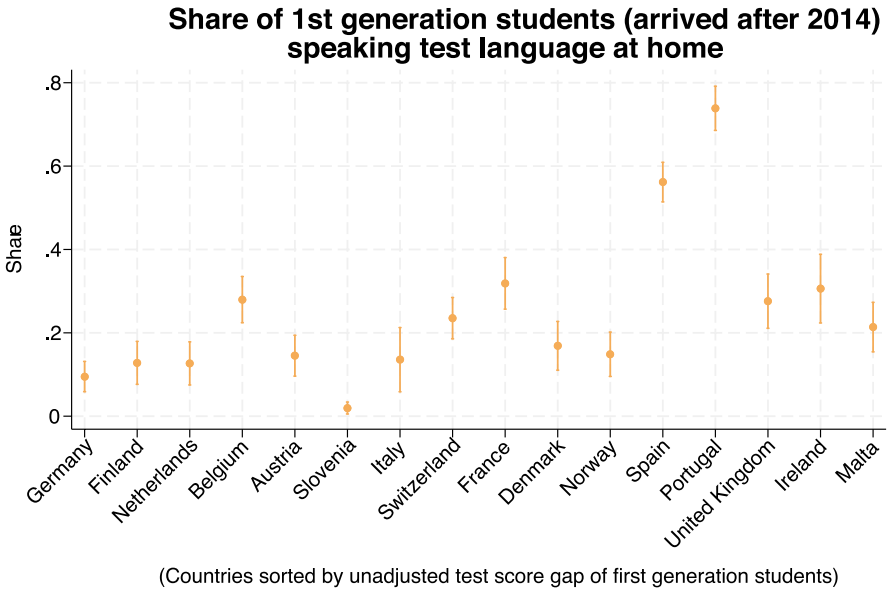

Fig. 4 Share of newly arrived immigrant students speaking test language at home, across countries. Source: PISA 2022, weighted analysis, our own calculations

proportion of first-generation immigrant students who arrived post-2014 who report speaking the language of the PISA test (i.e., the host country's language of instruction) at home. As with the previous figure, this graphic facilitates a straightforward cross-national comparison, illustrating how this dimension of linguistic integration varies across the European contexts under examination.

Subsequently, we calculate the adjusted test score gap between native students and the newly arrived (post-2014) immigrant students for each country (see Fig. 5). To achieve this, we specified a series of multiple regression models, again calculated independently for each nation. Within these models, the coefficient for the newly arrived dummy variable was estimated while statistically controlling for a vector of covariates: gender, age, ESCS, and the binary indicator for language spoken at home. This allows us to isolate the net performance gap, holding these critical individual- and family-level characteristics constant.

In the final, conclusive step, we leverage this multi-level design to test whether the academic performance of newly arrived immigrants is moderated by the institutional structure of the host country's educational system. Specifically, we investigate whether the magnitude of the performance gap is associated with the degree of educational stratification. Our rationale for using the *adjusted* test score gaps from Fig. 5 for this macro-level comparison is to mitigate confounding variables. By controlling for individual-level characteristics, we minimise the influence of compositional selection effects (i.e., differences in the *types* of immigrants) and other processes originating outside the school system, thereby isolating the institutional effect more effectively. To operationalise educational stratification, we employ the

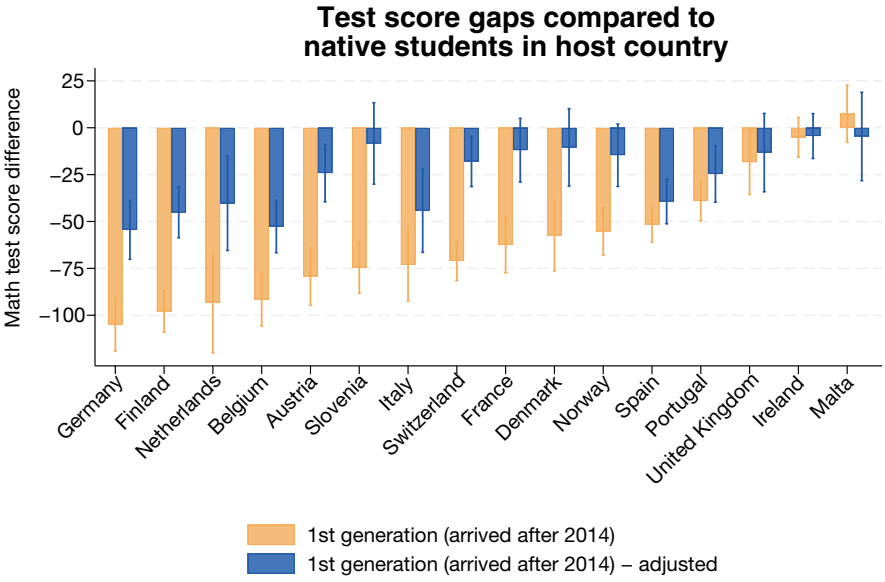

Fig. 5 Adjusted and unadjusted test score gaps from newly arrived students compared to native students, across countries. Source: PISA 2022, weighted analysis, our own calculations

established tracking index developed by Bol et al. (2014). This index quantifies the degree of differentiation (e.g., early tracking) within an education system; higher values denote a more highly stratified system. Among the countries in our sample, Germany (index value: 1.862) and Austria (index value: 1.817) exhibit the highest stratification scores, whereas Norway (index value: −1.043) and the United Kingdom (index value: −1.043) represent the systems with the lowest scores.

4 Results

Figure 1 serves to delineate the demographic composition of the final analytical sample across all 16 participating nations, specifically illustrating the relative proportions of native students, first-generation immigrant students, and second-generation immigrant students. Several key observations immediately emerge from this descriptive mapping. Firstly, with the singular and notable exception of Switzerland, native students constitute the clear numerical majority in all national contexts. The cross-national range is considerable: whereas in Finland, native students comprise 85% of the analytical sample, in Switzerland, this figure stands at only 44%. Secondly, a highly consistent pattern observed across all countries is that the relative share of second-generation students substantially exceeds the proportion of first-generation students. Switzerland exhibits the highest concentration of second-generation immigrants (44%), while Finland presents the lowest (10%). The share of first-generation students, conversely, is at its lowest in Italy (a mere 3% of the sample) and, once again, at its highest in Switzerland (15%). Thirdly, and perhaps most saliently for the subsequent comparative analyses, what is striking is the significant heterogeneity in these demographic compositions across the European landscape (compare, for instance, the immigrant population shares in Finland and the Netherlands with the high-prevalence contexts of Switzerland, Austria, and Germany).

Figure 2 visualises the unadjusted proficiency gaps in mathematical literacy. These gaps are plotted between native students (the reference category) and, respectively, second-generation immigrant students and first-generation immigrant students (the latter being further differentiated by their date of arrival, i.e., pre- or post-2014). It is immediately apparent, and in alignment with Hypothesis 1, that the majority of immigrant subgroups exhibit significantly lower academic performance relative to their native peers. The most profound finding, however, is the immense heterogeneity in the magnitude of these proficiency gaps across the different national contexts. While nations such as the United Kingdom, Ireland, and Malta exhibit negligible or, in most cases, statistically insignificant performance differentials, other countries—most notably Germany, the Netherlands, and Finland—reveal exceptionally large and substantively profound achievement gaps. In these three national contexts, the raw proficiency gap between newly arrived students (post-2014) and native-majority students approaches 100 PISA scale points, a disparity of substantial magnitude that is equivalent to approximately one full standard

deviation of the PISA test score distribution. Furthermore, it is evident that—as was theoretically anticipated and in accordance with Hypothesis 2—these performance gaps tend to attenuate across the generational sequence and with increased duration of residence. This pattern is, in some cases, quite dramatic. The results for Italy and Norway provide particularly lucid examples of this integration trajectory. Notably, the newly arrived students (post-2014) consistently exhibit the most pronounced performance disparities, a finding which justifies a more granular investigation of this specific subgroup in the subsequent steps, as well as a deeper examination into the macro-level factors driving this substantial cross-national divergence in observed outcomes.

As was articulated in the preceding theoretical section, one compelling explanation for these extensive country-level differences in performance gaps relates to differential socio-economic selectivity of the respective migrant cohorts. To empirically interrogate this compositional hypothesis, Fig. 3 plots the mean ESCS for both native students and the newly arrived (post-2014) immigrant subgroup. The findings confirm that, as expected, in the majority of host countries, newly arrived immigrant students originate from, on average, less advantaged socio-economic backgrounds than their native peers. It is, however, highly noteworthy that in Portugal, Ireland, and Malta, these two groups exhibit no statistically or substantively significant difference in their ESCS profiles. Moreover, in the United Kingdom, the socio-economic gap is markedly smaller than that observed in other nations. This pattern strongly suggests the presence of heterogeneous migration streams across Europe, implying fundamentally different modes of social incorporation and leading to varying degrees of social stratification. This interpretation is further bolstered by the observation that even among those countries that do exhibit substantial socio-economic gaps, the magnitude of this within-country differential varies considerably.

Figure 4 continues this exploration of compositional differences by delineating the proportion of newly arrived students who report speaking the PISA test language (i.e., the host country's primary language of instruction) within their domestic environment. Here again, a pronounced cross-national heterogeneity is immediately visible (compare, for instance, the vastly different situations in Slovenia and Portugal). Particularly conspicuous are the cases of Spain and Portugal, where a clear majority (well over 50%) of newly arrived immigrant students report speaking the test language at home. This finding almost certainly reflects migration streams from specific, linguistically proximate regions (e.g., South America in the case of Spain, and Lusophone Africa or Brazil in the case of Portugal). It is likewise discernible that in nations with major global languages (specifically, English and French), this proportion is comparatively higher than in the remaining European contexts. Collectively, these findings underscore that the initial endowment of newly arrived immigrant students with country-specific human capital—in this instance, linguistic proficiency—is highly variable across the host societies under investigation.

In Fig. 5, the unadjusted proficiency gaps are juxtaposed with the adjusted proficiency gaps for newly arrived students. To compute these adjusted gaps, we

specified a series of regression models that statistically control for a vector of key individual- and family-level covariates: student age, gender, ESCS, and the binary indicator for language spoken at home. It becomes unequivocally clear that, once these compositional factors are accounted for, the proficiency gaps in many countries diminish markedly (observe, for example, the substantial reductions in France, Denmark, and Slovenia). This demonstrates that these factors possess significant explanatory power in accounting for migrant-native performance disparities, thereby lending strong empirical support to Hypothesis 3. Upon examining the adjusted gaps, it is also apparent that the cross-national heterogeneity, while still present, is markedly reduced compared to that seen in the unadjusted models. The maximal range of the inter-country variance is reduced from the previous high of nearly 100 points to a mere 50 points. While this constitutes a significant reduction in heterogeneity, this remaining differential—equivalent to almost half a standard deviation—is still highly substantial. From a comparative sociological perspective, these results can be interpreted as evidence that different host nations must contend with vastly different types of immigrant cohorts, particularly in terms of their human capital and their social integration. These are selection-based processes that, for the most part, originate outside the direct purview of the formal education system.

The final analytical step, presented in Fig. 6, moves to the macro-level to test our institutional hypothesis. This figure plots the relationship between the adjusted test score gaps for newly arrived students (from Fig. 5) and the degree of institutional stratification within each country's education system, as operationalised by the tracking index (Bol et al., 2014). As data for the tracking index are not available for Malta, this country is necessarily excluded from this specific analysis. The resulting scatterplot indicates a positive association, corroborating the central premise of

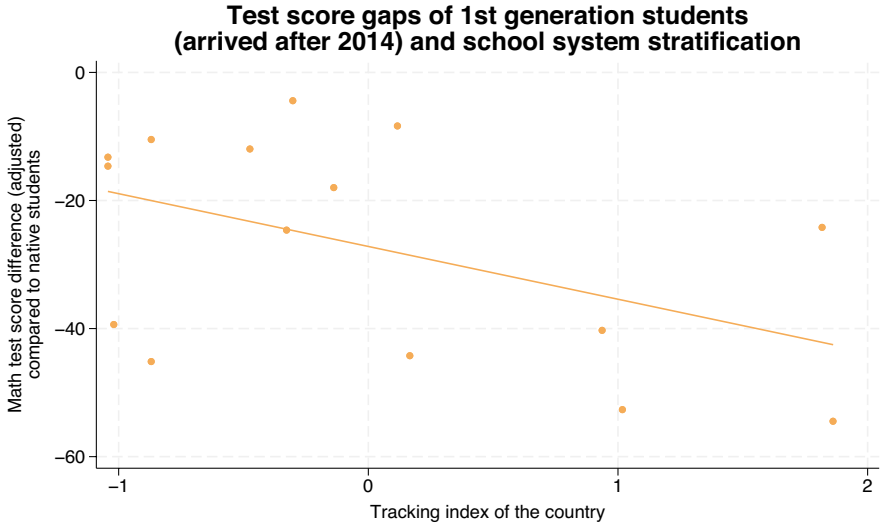

Fig. 6 Association between adjusted test score gaps and country-level educational stratification. Source: PISA 2022, weighted analysis and Bol et al. (2014), our own calculations

Hypothesis 4: namely, that countries with more highly stratified educational systems (i.e., higher tracking indices) are associated with larger adjusted proficiency gaps, even after accounting for individual-level differences in socio-economic status and home language. In summary, these analyses suggest that even when controlling for the compositional selectivity of the student populations, institutionally differentiated systems appear to exacerbate performance gaps between native and newly arrived students. However, a crucial caveat must be appended to this finding: the plot also reveals the existence of outlier countries which, despite maintaining nominally comprehensive or weakly stratified systems, nonetheless exhibit adjusted test score gaps of a comparable magnitude, suggesting the presence of more complex, multi-causal interactions.

5 Conclusion

This chapter embarked on a comparative analysis of the educational performance of newly arrived immigrant students—defined as those arriving in or after 2014—across 16 European host countries, using the PISA 2022 dataset. The empirical results confirm our baseline expectations (H1 and H2): immigrant-background students generally exhibit lower academic performance, and these disparities are by far the most pronounced for the newly arrived cohort. The unadjusted proficiency gaps for this group are alarmingly large in several countries, most notably Germany, the Netherlands, and Finland, where they approach 100 scale points. However, the most striking initial finding is the profound cross-national heterogeneity, with countries like the United Kingdom, Ireland, and Malta exhibiting statistically and substantively negligible gaps.

Our analysis demonstrates that this heterogeneity is, to a significant degree, a function of differential compositional selectivity (supporting H3). The socio-economic profile (ESCS) and linguistic capital (home language) of newly arrived students vary dramatically across Europe. In contexts like Portugal and Ireland, newly arrived students are, on average, socio-economically on par with their native peers. In contrast, in Germany and Austria, they are characterised by a markedly lower socio-economic profile. This confirms that host nations are contending with vastly different socio-demographic realities (Schmidt et al., 2021; Spörlein et al., 2020; Spörlein & Kristen, 2019). Indeed, once these compositional factors are statistically controlled, the proficiency gaps shrink considerably in most countries, and the cross-national variance is nearly halved. This finding underscores that a large portion of the ethnic penalty is, in fact, a manifestation of socio-economic inequality and a lack of host-country-specific linguistic capital.

Yet, even after accounting for these crucial compositional effects, substantial net gaps persist, and these gaps vary systematically across nations. Our final analysis

lends support to our institutional hypothesis (H4). We find a positive association between the degree of educational stratification (Bol et al., 2014) and the magnitude of the adjusted test score gap. This suggests that highly tracked systems, such as those in Germany and Austria, are less effective at mitigating the initial disparities of newly arrived students. By sorting students early and rigidly, these systems appear to amplify, rather than mitigate, imported inequalities, a finding that aligns with a significant body of comparative stratification research (Terrin & Triventi, 2022; Triventi et al., 2016). The consequences of such systemic failures are well-documented, leading to pronounced inequalities in subsequent educational transitions and, ultimately, to less favourable positions in the labour market (Meyer & Winkler, 2025; Nennstiel, 2021, 2022a).

Several limitations must qualify these conclusions. First, the PISA dataset is cross-sectional. While our findings are robustly associative, they cannot, and do not, imply direct causality (Rutkowski et al., 2024). We are observing students at a single point in time, and we cannot dismiss the role of unobserved prior knowledge or differential schooling trajectories. Second, and most critically for the theme of this volume, the PISA data do not permit a direct identification of refugee status. Our "post-2014" cohort is a proxy intended to capture this recent wave of migration, but it necessarily aggregates forced migrants (Kogan & Kalter, 2020) with labour migrants and students arriving via family reunification. As national-level studies focusing specifically on refugee children in Germany have often documented even more pronounced disparities (Glinka & Winkler, 2024; Olczyk et al., 2025; Schipolowski et al., 2021), it is highly probable that our results underestimate the true magnitude of educational inequalities facing this specific group. Finally, data limitations prevented an analysis of school- or classroom-level compositional effects (Glinka & Winkler, 2024; Seuring et al., 2020), which remains a critical avenue for future research. A third methodological constraint pertains to the use of the ESCS index as a composite measure of socioeconomic background (Avvisati, 2020). While this index effectively summarises economic, social, and cultural status, it risks conflating distinct dimensions that frequently diverge within immigrant populations. Research highlights that first-generation immigrants are often positively selected from the upper social strata of their countries of origin yet face significant human capital devaluation upon arrival (Schmidt et al., 2021; Spörlein et al., 2020; Spörlein & Kristen, 2019). Consequently, many newly arrived families may possess high levels of formal education that are not yet reflected in their occupational status or material wealth in the host country. This discrepancy implies that the aggregate ESCS measure may obscure important nuances regarding the educational resources these families actually possess. Future research would therefore benefit from disaggregating these components to disentangle the specific influence of parental education from realised socioeconomic status in the destination context.

A further limitation concerns the unavailability of classroom-level identifiers in PISA, which precludes an analysis of immediate peer composition. This restriction

is particularly significant given the variation in national migration regimes and destination attractiveness across Europe. Consequently, we could not account for the possibility that in countries with lower immigration rates, newly arrived students may enter less segregated classroom environments. This context could potentially facilitate faster language acquisition and distinct achievement trajectories that our analysis remains unable to capture.

Finally, it is necessary to reflect on the conceptual premise of comparing newly arrived students with native-born peers. As noted by Schipolowski et al. (2021), such comparisons entail an inherent structural asymmetry that can be perceived as unfair. Given that newly arrived adolescents have had limited exposure to the host country's educational environment and often carry histories of interrupted schooling or forced migration, lower average performance is to some extent a predictable outcome of these initial disparities. However, despite these unequal starting conditions, such comparative analyses remain sociologically indispensable. Since educational credentials function as central gatekeepers for future life chances and labour market access, documenting achievement gaps is essential to reveal the actual stratification risks facing this population. Moreover, placing these disparities within a cross-national framework allows us to identify which institutional systems succeed in mitigating these profound initial disadvantages, thereby offering evidence for how to facilitate more equitable integration trajectories.

Given the pivotal role of linguistic proficiency in mediating mathematical competence (Schipolowski et al., 2021), measures that support the acquisition of the language of instruction appear to be particularly important. In this context, recent evidence suggests that administrative conditions may influence skill development; for instance, prolonged waiting periods for school enrolment and insecure residence status have been associated with lower destination-language skills (Winkler & Carwehl, 2025). Furthermore, the specific design of initial reception measures could be a contributing factor, as extended attendance in preparatory classes has been linked to persistent linguistic disadvantages compared to models facilitating earlier immersion (Winkler & Carwehl, 2025). Consequently, efforts to mitigate achievement gaps might benefit from considering not only the effects of curricular tracking but also the potential impact of administrative delays and the organisation of the transition into mainstream education.

In conclusion, this chapter provides a comprehensive, comparative overview of the educational integration of newly arrived students in the post-2014 era. Our findings send a dual message to policymakers. On the one hand, interventions must be targeted at mitigating the profound socio-economic and linguistic barriers that many, though not all, newly arrived students encounter. On the other hand, policy must look inward, at the institutional structures that shape opportunity. Education systems that prioritise early, rigid stratification may be systematically failing their most vulnerable new members.

Competing Interests The author declares no competing interests.

Appendix

Table 1 Valid cases for individual model variables and complete cases included in the analysis, by country

	Gender	Test scores	Age	Migrant generation	Language spoken	ESCS	Parents' Birthplace	Complete cases
Austria	6151	6151	6151	6039	6090	5981	6032	5927
Belgium	8286	8286	8286	7998	8114	8000	8009	7807
Denmark	6200	6200	6200	5936	5964	5908	5921	5846
Finland	10,239	10,239	10,239	10,004	10,031	9931	9965	9818
France	6770	6770	6770	6541	6643	6572	6521	6348
Germany	6116	6116	6116	5427	5463	5314	5418	5267
Ireland	5569	5569	5569	5469	5533	5520	5496	5436
Italy	10,552	10,552	10,552	10,345	10,473	10,451	10,343	10,277
Malta	3127	3127	3127	2930	2978	2963	2920	2879
Netherlands	5046	5046	5046	4769	4880	4860	4777	4662
Norway	6611	6611	6611	6164	6252	6210	6130	6042
Portugal	6793	6793	6793	6650	6702	6674	6653	6605
Slovenia	6721	6721	6721	6555	6599	6592	6542	6499
Spain	30,800	30,800	30,800	29,366	29,818	29,648	29,320	29,059
Switzerland	6829	6829	6829	6701	6758	6596	6700	6494
United Kingdom	12,972	12,972	12,972	11,167	11,401	11,084	11,196	10,761
Total	138,782	138,782	138,782	132,061	133,699	132,304	131,943	129,727

References

Avvisati, F., & Keslair, F. (2014). REPEST: Stata module to run estimations with weighted replicate samples and plausible values. Statistical Software Components, Boston College Department of Economics. https://ideas.repec.org/c/boc/bocode/s457918.html

Avvisati, F. (2020). The measure of socio-economic status in PISA: A review and some suggested improvements. *Large-scale Assessments in Education, 8*(1), 8. https://doi.org/10.1186/s40536-020-00086-x

Bol, T., Witschge, J., Van de Werfhorst, H., & Dronkers, J. (2014). Curricular tracking and central examinations: Counterbalancing the impact of social background on student achievement in 36 countries. *Social Forces, 92*(4), 1545–1572. https://doi.org/10.1093/sf/sou003

Drouhot, L. G., & Nee, V. (2019). Assimilation and the second generation in Europe and America: Blending and segregating social dynamics between immigrants and natives. *Annual Review of Sociology, 45*(1), 177–199. https://doi.org/10.1146/annurev-soc-073117-041335

Fleischmann, F., & Kristen, C. (2014). Gender inequalities in the education of the second generation in western countries. *Sociology of Education, 87*(3), 143–170. https://doi.org/10.1177/0038040714537836

Gamoran, A. (1987). The stratification of high school learning opportunities. *Sociology of Education, 60*(3), 135–155.

Glinka, H., & Winkler, O. (2024). Mathematisch-naturwissenschaftliche Kompetenzen zugewanderter Heranwachsender mit und ohne Fluchtbiografie: Welche Rolle spielt die sozioökonomische und ethno-linguale Schulklassenkomposition? *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie, 76*(1), 57–88. https://doi.org/10.1007/s11577-024-00938-2

Gutfleisch, T., & Nennstiel, R. (2025). Bridging the gap: Gender-specific preferences in STEM occupations in vocational education and training. *Empirical Research in Vocational Education and Training, 17*, 12. https://doi.org/10.1186/s40461-025-00187-3

Heath, A. F., Rothon, C., & Kilpi, E. (2008). The second generation in Western Europe: Education, unemployment, and occupational attainment. *Annual Review of Sociology, 34*(1), 211–235. https://doi.org/10.1146/annurev.soc.34.040507.134728

Heath, A. F. (Ed.). (2014). *Unequal attainments: Ethnic educational inequalities in ten Western countries*. Oxford University Press.

Heisig, J. P., & Matthewes, S. H. (2022). No evidence that strict educational tracking improves student performance through classroom homogeneity: A critical reanalysis of Esser and Seuring (2020). *Zeitschrift für Soziologie, 51*(1), 99–111. https://doi.org/10.1515/zfsoz-2022-0001

Kogan, I., & Kalter, F. (2020). An empirical--analytical approach to the study of recent refugee migrants in Germany. *Soziale Welt, 71*(1/2), 3–23.

Laubenthal, B. (2023). Introduction: Assimilation, integration or transnationalism? An overview of theories of migrant incorporation. *International Migration, 61*(1), 84–91. https://doi.org/10.1111/imig.13118

Meyer, F., & Winkler, O. (2025). Refugee men and women in the segmented VET market in Germany: On their career prospects and the role of resource inequality. *Empirical Research in Vocational Education and Training, 17*(1), 22. https://doi.org/10.1186/s40461-025-00197-1

Nennstiel, R. (2021). Do more demanding lower secondary school certificates for minority students pay off? A comparison of VET access between Germany and German-speaking Switzerland. *SozW Soziale Welt, 72*(3), 313–342. https://doi.org/10.5771/0038-6073-2021-3-313

Nennstiel, R. (2022a). Explaining ethnic differences in access to VET in German-speaking Switzerland: The effects of application behavior and resources. *Zeitschrift für Soziologie, 50*(6), 396–414. https://doi.org/10.1515/zfsoz-2021-0024

Nennstiel, R. (2022b). No Matthew effects and stable SES gaps in math and language achievement growth throughout schooling: Evidence from Germany. *European Sociological Review, 38*(6), jcac062. https://doi.org/10.1093/esr/jcac062

Nennstiel, R. (2022c). Positive ethnic choice effects are persistent over time but not across minority generations: Evidence from Switzerland. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie, 74*(4), 499–524. https://doi.org/10.1007/s11577-022-00865-0

Nennstiel, R., & Gilgen, S. (2024). Does chubby can get lower grades than skinny Sophie? Using an intersectional approach to uncover grading bias in German secondary schools. *PLoS One, 19*(7), e0305703. https://doi.org/10.1371/journal.pone.0305703

OECD. (2023a). *PISA 2022 results (volume I)*. OECD Publishing.

OECD. (2023b). *PISA 2022 database*. https://www.oecd.org/en/data/datasets/pisa-2022-database.html

Olczyk, M., Heppt, B., Winkler, O., Schipolowski, S., & Edele, A. (2025). Academic achievement of refugee students at the end of primary school in Germany. *Zeitschrift für Erziehungswissenschaft, 28*(3), 557–579. https://doi.org/10.1007/s11618-025-01315-1

Penner, A. M. (2008). Gender differences in extreme mathematical achievement: An international perspective on biological and social factors. *American Journal of Sociology, 114*(S1), S138–S170. https://doi.org/10.1086/589252

Pokropek, A., Borgonovi, F., & Jakubowski, M. (2015). Socio-economic disparities in academic achievement: A comparative analysis of mechanisms and pathways. *Learning and Individual Differences, 42*, 10–18.

Ritchie, S. J., & Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological Science, 24*(7), 1301–1308. https://doi.org/10.1177/0956797612466268

Rutkowski, D., Rutkowski, L., Thompson, G., & Canbolat, Y. (2024). The limits of inference: reassessing causality in international assessments. *Large-scale Assessments in Education, 12*(1), 9. https://doi.org/10.1186/s40536-024-00197-9

Schipolowski, S., Edele, A., Mahler, N., & Stanat, P. (2021). Mathematics and science proficiency of young refugees in secondary schools in Germany. *Journal for Educational Research Online, 13*(1), 78–104.

Schmidt, R., Kristen, C., & Mühlau, P. (2021). Educational selectivity and immigrants' labour market performance in Europe. *European Sociological Review, 38*(2), 252–268. https://doi.org/10.1093/esr/jcab042

Seuring, J., Rjosk, C., & Stanat, P. (2020). Ethnic classroom composition and minority language use among classmates: Do peers matter for students' language achievement? *European Sociological Review, 36*(6), 920–936. https://doi.org/10.1093/esr/jcaa022

Spörlein, C., & Kristen, C. (2019). Why we should care about regional origins: Educational selectivity among refugees and labor migrants in Western Europe. *Frontiers in Sociology, 4*, 39. https://doi.org/10.3389/fsoc.2019.00039

Spörlein, C., Kristen, C., Schmidt, R., & Welker, J. (2020). Selectivity profiles of recently arrived refugees and labour migrants in Germany. *SozW Soziale Welt, 71*(1-2), 54–89.

Terrin, É., & Triventi, M. (2022). The effect of school tracking on student achievement and inequality: A meta-analysis. *Review of Educational Research, 93*(2). https://doi.org/10.3102/00346543221100850

Triventi, M., Kulic, N., Skopek, J., & Blossfeld, H. P. (2016). Secondary school systems and inequality of educational opportunity in contemporary societies. In H. P. Blossfeld et al. (Eds.), *Models of secondary education and social inequality: An international comparison* (pp. 3–24). Edward Elgar Publishing.

Triventi, M., Vlach, E., & Pini, E. (2022). Understanding why immigrant children underperform: Evidence from Italian compulsory education. *Journal of Ethnic and Migration Studies, 48*(10), 2324–2346.

Urruticoechea, A., Oliveri, A., Vernazza, E., Giménez-Dasí, M., Martínez-Arias, R., & Martín-Babarro, J. (2021). The relative age effects in educational development: A systematic review. *International Journal of Environmental Research and Public Health, 18*(17), 8966. https://doi.org/10.3390/ijerph18178966

Winkler, O., & Carwehl, A.-K. (2025). Institutional conditions and acquisition of language skills among young refugees: Investigating the German context. *Acta Sociologica*. https://doi.org/10.1177/00016993251351531