ROCKWOOL Foundation Berlin
Institute for the Economy and the Future of Work (RFBerlin)

DISCUSSION PAPER SERIES 048/26

Host Country Language and School Integration of Immigrant Students

Yu Qin, Michael Vlassopoulos, Jackline Wahba

www.rfberlin.com FEBRUARY 2026

Host Country Language and School Integration of Immigrant Students

Authors

Yu Qin, Michael Vlassopoulos, Jackline Wahba

Reference

JEL Codes: I21, J15, I31

Keywords: Language proficiency, immigrant integration, bullying, school belonging, PISA, instrumental variables

Recommended Citation: Yu Qin, Michael Vlassopoulos, Jackline Wahba (2026): Host Country Language and School Integration of Immigrant Students. RFBerlin Discussion Paper No. 048/26

Imprint

RFBerlin
ROCKWOOL Foundation Berlin –
Institute for the Economy
and the Future of Work

Gormannstrasse 22, 10119 Berlin
Tel: +49 (0) 151 143 444 67
E-mail: info@rfberlin.com
Web: www.rfberlin.com

Host Country Language and School Integration of Immigrant Students

Yu Qin Michael Vlassopoulos Jackline Wahba

February 2026

Abstract

Language proficiency is a key determinant of immigrant integration. This paper examines the causal impact of host-country language proficiency (proxied by reading test scores) on school integration and bullying among first-generation immigrant students across 16 OECD destination countries, using data from the 2015, 2018, and 2022 waves of the Programme for International Student Assessment (PISA). We employ an instrumental variable strategy exploiting exogenous variation in exposure to the host-country language, measured through the interaction of immigrants' age at arrival and the linguistic distance between their mother tongue and the host-country language. We find that a one-standard-deviation increase in reading proficiency raises a standardized index of school integration by about 0.56 standard deviations, and reduces a standardized bullying index by about 0.59 standard deviations. The protective effect against bullying is stronger for boys, whereas integration gains are more similar across genders. We also find positive effects on academic achievement and grade progression, and links to more ambitious expectations, better teacher relations, and a stronger non-cognitive profile that plausibly explains the integration effects. These results suggest the importance of language proficiency as an input into the joint production of cognitive and psychosocial outcomes for immigrant youth.

Keywords: Language proficiency, immigrant integration, bullying, school belonging, PISA, instrumental variables

JEL Codes: I21, J15, I31

1 Introduction

Language proficiency is a key determinant of immigrant integration. For immigrant children, proficiency in the host-country language is central for educational achievement and social participation. It enables access to the curriculum, effective communication with teachers and peers, and active engagement in school activities. In contrast, limited language skills constrain learning, impede interaction, and reinforce educational disparities. In 2020, about 19% of young people under the age of 16 were of immigrant backgrounds in OECD countries (OECD, 2025). Yet, according to PISA 2022, only about half of 15-year-old immigrant students spoke the host-country language at home–45% among first-generation and 47% among second-generation students (OECD, 2025). This substantial variation in language exposure raises a natural empirical question: how does proficiency in the host-country language causally affect school integration for immigrant children? Examining this question is critical for understanding human capital formation and for designing policies that promote both educational success and social inclusion.

Language proficiency is a central component of human capital. A large body of research has documented its role in shaping immigrants' economic integration in the labor market (e.g. Chiswick and Miller, 1995; Bleakley and Chin, 2004; Dustmann and van Soest, 2001; Schmid, 2023), as well as its effects on a range of other outcomes, including health, fertility, and social and political participation (Bleakley and Chin, 2010; Clarke and Isphording, 2017). Beyond these economic returns for adult immigrants, a smaller literature has examined the role of language proficiency in the educational outcomes of immigrant children. This literature finds persistent achievement gaps between immigrant and native-born students and attributes a substantial share of these disparities to differences in host-country language proficiency (e.g. Schnepf, 2007; Dustmann et al., 2012; Entorf, 2015; Geay et al., 2013; OECD, 2018). These gaps are particularly consequential because language skills acquired early in life are a key input into human capital formation and, as such, a fundamental determinant of later educational and labor market success.

Despite this extensive evidence on the economic returns to language proficiency, much less attention has been paid to its role in shaping the social and emotional dimensions of immigrant children's schooling, and in particular their school integration, including sense of belonging, peer relationships, and exposure to bullying. Yet these dimensions are likely to matter for long-run outcomes, as they affect engagement with school, mental well-being, and the accumulation of both cognitive and non-cognitive skills. This paper addresses this gap by providing causal evidence on the effects of host-country language proficiency on multiple dimensions of school integration and well-being for first-generation immigrant students.

We use data from the Programme for International Student Assessment (PISA), covering about 14,000 first-generation immigrant students across 2,620 schools in 16 OECD destination countries. The richness and comparability of PISA data allow us to study integration and bullying measures across diverse institutional contexts. We use the PISA reading test score as a measure of functional proficiency in the host-country language. As a robustness check, we additionally exploit information on the language spoken at home.

To address endogeneity and measurement error in reading test scores, we employ an instrumental variable (IV) strategy exploiting predicted difficulty of language acquisition based on immigrants' age at arrival and the linguistic distance between their mother tongue and the host-

country language (Bleakley and Chin, 2004, 2010; Isphording et al., 2016). The idea is that children who arrive earlier and come from linguistically closer origins are exposed to the host-country language for longer and can acquire it at lower cost, thus creating quasi-experimental variation in the costs and intensity of language acquisition, conditional on school fixed effects and controls. Specifically, we instrument reading proficiency with the interaction between age at arrival and linguistic distance, while controlling flexibly for the main effects of arrival age and linguistic distance. Using this strategy, we provide causal evidence that host country language proficiency substantially improves school integration and reduces bullying among first-generation immigrant students.

We document four main results. First, we find significant effects on integration: a one standard deviation increase in reading proficiency raises the integration index by 0.56 standard deviations in 2SLS estimates, compared to 0.065 in OLS, indicating substantial attenuation bias in naive estimates due to measurement error and endogeneity. Second, language proficiency reduces bullying exposure by 0.59 standard deviations, indicating a strong protective effect that OLS substantially underestimates. Third, the protective effects against bullying are particularly pronounced for male students, while integration benefits are more evenly distributed. Finally, we present evidence that immigrant students' language proficiency is associated with improvements in native peers' outcomes, a pattern consistent with peer spillovers.

We complement these main results with evidence on secondary outcomes and possible channels. Using the same IV strategy, we show that higher reading proficiency improves immigrant students' performance in mathematics and science and reduces the probability of having repeated a grade, with no precisely estimated effect on absenteeism. We also find positive effects on expectations about education and occupation suggesting an important role for language proficiency in shaping aspirations.

Host-country language proficiency can plausibly affect school integration and bullying through three linked channels. First, better language skills reduce communication frictions with peers, making it easier to follow conversations, understand social norms, and form friendships rather than being sidelined. Second, language proficiency improves communication with teachers and access to instruction, which can increase support and enhance engagement. Third, language is likely to be intertwined with a broader set of socio-emotional or non-cognitive skills – such as resilience, empathy, and growth mindset – that help students manage social interactions and cope with setbacks. Using additional PISA modules, we document positive associations between reading proficiency and educational and occupational expectations, teacher–student relationship quality, and a range of socio-emotional skills and psychological traits. While these associations are descriptive, they are consistent with the idea that language proficiency affects school integration and bullying through improved communication with peers and teachers, higher aspirations, better teacher support, and stronger socio-emotional capacities.

The paper contributes to three strands of research. First, it contributes to the literature on language as an input into human capital formation among immigrant children. A few studies show that reading and language proficiency play a critical role in shaping academic achievement among immigrant students, particularly in mathematics (Isphording et al., 2016; Cavallo and Russo, 2025; Ortega and Ludwig, 2023). For example, Isphording et al. (2016) estimates the causal effect of reading proficiency on mathematics outcomes among immigrant students. More

recently, Tumen et al. (2025) evaluate a large-scale Turkish language training programme for Syrian refugee children, finding improvements in Turkish and math scores and reductions in school absences, using administrative data and a staggered difference-in-differences design. While these studies establish language proficiency's role in cognitive skill development, they do not examine whether, and to what extent, language proficiency shapes the social and emotional dimensions of schooling that independently predict long-term well-being and labor market success.

Second, another related strand of the literature has examined the assimilation of refugee children, in particular, focusing on their determinants of language acquisition. Boucher et al. (2021) estimate the causal effect of inter-ethnic exposure in early childhood programmes on the social integration of refugee children in Türkiye. Alan et al. (2021) analyze how ethnic segregation in Turkish schools affects peer violence, social integration, and Turkish language acquisition. Schilling and Höckel (2025) analyze the academic consequences of separate preparatory language classes. We add to this literature by demonstrating the importance of language proficiency itself for school integration and well-being in a multi-country setting, beyond its effects on academic outcomes.

Third, our paper relates to the literature on bullying and peer victimization. A small number of economic studies document the adverse consequences of bullying for human capital formation. For example, Sarzosa (2024) and Eriksen et al. (2014) show that bullying depletes cognitive and non-cognitive skills during childhood and harms subsequent school performance, while Brown and Taylor (2008) finds that the negative effects of bullying on educational attainment persist into adulthood. However, despite this evidence on consequences, the economic literature has paid relatively little attention to the determinants of bullying and peer victimization.

A broader psychology and public-health literature suggests that language barriers are a salient risk factor for bullying among immigrant youth, as limited proficiency in the host-country language may reinforce social exclusion, impede peer communication, and increase vulnerability to peer aggression. For instance, a systematic review by Pottie and Hassan (2015) concludes that first-generation immigrant adolescents from non-native-language-speaking backgrounds are more likely to experience bullying and peer victimization than native-born youth. Similarly, Maynard et al. (2016) document higher rates of bullying victimization among immigrant youth relative to their native-born peers in the United States. While informative, this literature is largely descriptive and does not establish whether language proficiency itself causally reduces bullying.

Recent economic work has begun to move in this direction. Tumen et al. (2025) provides causal evidence that reductions in language barriers enhance academic performance and school participation, while also improving mental health outcomes, including lower anxiety and reduced exposure to bullying. Our paper builds on this insight by moving beyond program-based evidence to estimate the causal effects of language proficiency as a continuous skill on bullying and school integration across a wide set of OECD destination countries.

The analysis in this paper is motivated by the critical role of language proficiency in immigrant students' integration. Our findings show that host country language proficiency shapes not only academic achievement but also the social and emotional dimensions of schooling. By establishing its causal impact on school integration, the paper helps explain persistent inequalities in human capital formation and informs policies aimed at the successful inclusion of immigrant students.

The rest of the paper is organized as follows. Section 2 describes the PISA data, the con-

struction of the outcome and language proficiency measures. Section 3 sets out the empirical strategy, while Section 4 presents the main results. Section 5 reports robustness checks while Section 6 discusses the broader implications and concludes.

2 Data and Measures

We draw on data from the Programme for International Student Assessment (PISA), a triennial global assessment administered by the Organisation for Economic Co-operation and Development (OECD) to nationally representative samples of 15-year-olds. PISA measures students' competencies in mathematics, reading, and science in more than 100 participating countries.[1]. Our analysis uses the 2015, 2018, and 2022 waves of PISA, which are the only rounds that simultaneously include detailed information on students' socio-emotional well-being at school, bullying experiences, and migration background.

To measure school integration outcomes, we use a set of PISA questions capturing students' sense of belonging and peer relationships. Specifically, students report whether they feel like an outsider, feel they belong at school, feel lonely, feel awkward at school, make friends easily, and believe that other students like them. We additionally exploit a series of questions on bullying experiences, including whether other students deliberately excluded them, made fun of them, took or damaged their belongings, spread unpleasant rumors about them, threatened them, or hit or pushed them (see Table A1). From these items, we construct two standardized indices: an integration index, capturing socio-emotional inclusion at school, and a bullying index, capturing exposure to peer victimization. Each index is formed by averaging the relevant components and standardizing the resulting measure to mean zero and unit variance.

Our main variable of interest is language proficiency, proxied by students' performance in the PISA reading literacy assessment. Reading literacy captures the ability to understand, interpret, and use written language across a range of contexts and therefore provides a comprehensive measure of functional language competence. Importantly, the PISA reading assessment focuses on higher-order comprehension and information processing skills rather than mechanical decoding, making it particularly suitable for capturing language proficiency among students educated in a non-native language. The reading assessment provides ten plausible values per student, which represent random draws from the posterior distribution of their latent reading ability and account for measurement uncertainty in large-scale assessments.Following OECD guidelines, descriptive statistics are computed using the main sampling weight together with 80 replicate weights and Fay's Balanced Repeated Replication (BRR) to obtain population-representative estimates. In the regression analyses, all models are estimated separately for each plausible value, and coefficients and standard errors are combined using Rubin's multiple-imputation rules. For ease of interpretation and cross-country comparability, the reading proficiency scores are standardized to have a mean of zero and standard deviation one within each destination country and survey wave. This standardized measure is used throughout the empirical analysis (For more details see Section A.1).

As an alternative and complementary proxy for language proficiency, we also exploit in-

[1]See https://www.oecd.org/pisa/data/

formation on language spoken at home. PISA asks students whether they usually speak the host-country language at home. We construct a binary indicator equal to one if the student reports speaking the host country language at home, and zero otherwise. This measure captures exposure to the destination language in the family environment and reflects an important dimension of language acquisition that is plausibly orthogonal to school-based learning. We use this indicator in robustness checks to assess whether results are sensitive to alternative definitions of language proficiency.

PISA collects detailed information on students' migration backgrounds, including students' own country of birth and the birth countries of their parents. Students who report being born abroad are classified as first-generation immigrants. When students' country of birth is missing, we use parental information: we assign the father's country of birth as the country of origin, and if this is unavailable, the mother's country of birth. Using this information, we construct origin–destination pairs, where the destination is the PISA test country and the origin is the reported (or imputed) country of birth.

To ensure representativeness and sufficient variation for identification, we restrict the sample to (i) destination countries where immigrant students account for at least 5 percent of the student population; (ii) origin countries with at least 100 immigrant student observations; and (iii) origin–destination pairs with a minimum of ten observations. These restrictions yield a final sample of 16 destination countries, 51 origin countries, and 2,620 schools, covering 14,118 first-generation immigrant students. Table A2 reports the distribution of origin and destination countries.

Table 1 presents descriptive statistics for the main variables used in the analysis, separately for native students and first-generation immigrant students. Table A1 provides detailed variable definitions, while Table A3 reports descriptive statistics for categorical variables. PISA provides rich information on student and school characteristics, which we use as controls in the empirical analysis. Missing values in individual- and school-level characteristics are imputed using the median within the corresponding level. The extent of missingness across variables is summarized in Table A4. Table 1 Panel A shows that immigrant students exhibit, on average, lower levels of school integration and higher exposure to bullying than native students, with both differences being statistically significant. At the same time, immigrant students underperform academically, with lower mathematics and science scores, and are more likely to have repeated a grade. Absenteeism rates differ only modestly across groups, although the difference is statistically significant. Panel B documents pronounced gaps in language proficiency. Immigrant students score substantially lower in reading than native students, with a difference exceeding 0.2 standard deviations. Moreover, only about half of immigrant students report speaking the host-country language at home, highlighting large heterogeneity in exposure to the destination language outside school. Panel C reports student-level characteristics. While age and gender distributions are broadly similar across groups, immigrant students are, on average, enrolled in slightly lower grades. Within the immigrant sample, there is substantial variation in age at arrival and linguistic distance between the mother tongue and the host-country language, which underpin our instrumental-variable strategy, as well as in cultural distance, which is used as an

additional control in robustness checks.[2] Finally, Panel D reports school-level characteristics, showing that immigrant students are slightly more likely to attend public schools and schools with higher average reading performance and larger immigrant student shares.

Overall, the table illustrates large and systematic disparities between native and immigrant students in language proficiency, school integration, exposure to bullying, and academic outcomes. These patterns underscore the potential role of host-country language proficiency as a central input into both cognitive and psychosocial dimensions of immigrant integration, motivating the causal analysis that follows. Appendix A provides variable definitions, origin–destination composition, categorical breakdowns, and missingness patterns for the analysis sample.

Table 1: **Descriptive Statistics by Migration Status**

	Native students		**Immigrant students**		**Group difference**	
Variable	Mean	SD	Mean	SD	(Native – Immigrant)	t-statistic
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Integration outcomes and academic performance						
Integration	0.028	0.995	-0.094	1.009	0.122	14.140***
Bullying	-0.019	0.967	0.038	1.107	-0.057	-6.579***
Math score	0.093	0.976	-0.061	1.066	0.154	18.203***
Science score	0.102	0.977	-0.106	1.060	0.209	24.626***
Repeat grade	0.093	0.290	0.165	0.372	-0.072	-28.337***
Absenteeism	0.266	0.442	0.254	0.432	0.012	2.508**
Panel B: Language proficiency						
Reading score	0.098	0.972	-0.116	1.071	0.214	25.236***
Speaking host language at home	–	–	0.505	0.500	–	–
Panel C: Student level						
Age	15.895	0.307	15.855	0.352	0.040	14.894***
Grade	9.786	0.694	9.758	0.860	0.027	4.512***
Female	0.496	0.500	0.486	0.500	0.010	2.228**
Arrival age	–	–	7.918	4.783	–	–
Linguistic distance	–	–	0.644	0.289	–	–
Cultural distance	–	–	0.111	0.073	–	–
N (students)	259,699		14,118			
Panel D: School level						
Public school	0.795	0.404	0.809	0.393	-0.014	-0.890
Share of School Immigration	0.108	0.092	0.204	0.138	-0.096	-40.994***
School-average immigrants' reading score	-0.227	0.840	-0.208	0.749	-0.019	-1.023
School-average immigrant arrival age	7.227	3.921	7.852	3.200	-0.625	-7.426***
N (schools)	8,377		2,620			

Notes: Columns (1) and (2) report means and standard deviations for native students, and Columns (3) and (4) for first-generation immigrant students. Column (5) reports differences in means, and Column (6) the associated t-statistics. Reading, mathematics, and science scores are standardized z-scores constructed within country and survey wave using PISA plausible values. Integration is a standardized index of school belonging, and bullying a standardized index of bullying exposure. Bullying, grade repetition, and absenteeism are subject to item non-response, so sample sizes vary (13,448; 14,032; and 9,172, respectively). Panels A–C report student-level variables using all available observations (259,699 native students; 14,118 immigrant students), and Panel D reports school-level variables based on distinct school identifiers (8,377 and 2,620 schools, respectively). *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Source: PISA 2015, 2018, 2022; CEPII; WVS.

[2]We obtain country-pair measures of cultural distance from Geopolitical Distance (`https://www.geopoliticaldistance.org/cultural-distance`); these indices are constructed using the World Values Survey (WVS).

3 Empirical Strategy

We study how host-country language proficiency affects first-generation immigrant students' school integration and exposure to bullying. Our unit of observation is student i in school s, destination country c, and PISA wave t (2015, 2018, 2022). Outcomes and reading are standardized within country and wave.

We write the outcome equation in terms of latent language proficiency L_{isct},

$$Y_{isct} = \beta\, L_{isct} + X'_{isct}\gamma + \phi_s + \lambda_t + \varepsilon_{isct}, \tag{1}$$

but only observe the proxy R_{isct} (PISA reading test score). We therefore estimate

$$Y_{isct} = \beta\, R_{isct} + X'_{isct}\gamma + \phi_s + \lambda_t + \zeta_{isct}, \tag{2}$$

where X_{isct} contains student and school characteristics and ϕ_s are school fixed effects. Standard errors are clustered at the school level. Because $R_{isct} = L_{isct} + u_{isct}$ is measured with error and is endogenous to unobservables, OLS is biased and attenuated.

Instrumental Variables Following Bleakley and Chin (2004, 2010); Isphording et al. (2016), we instrument reading with the interaction of arrival age and linguistic distance:

$$Z_{isct} \;=\; \text{ArrivalAge}_{isct} \times \text{LingDist}_{isct}.$$

The identifying idea is that arriving later reduces host-country language proficiency more strongly when the origin language is more distant from the host language. In all specifications we control directly for LingDist_{isct} and flexibly for ArrivalAge_{isct}, as well as the full vector of covariates and fixed effects, so identification comes from the differential impact of later arrival on reading proficiency across more versus less linguistically distant origin languages.

The first stage is

$$R_{isct} \;=\; \pi\, Z_{isct} \;+\; \delta_1\, f(\text{ArrivalAge}_{isct}) \;+\; \delta_2\, \text{LingDist}_{isct} \;+\; X'_{isct}\Pi \;+\; \phi_s \;+\; \lambda_t \;+\; \eta_{isct}, \tag{3}$$

where $f(\text{ArrivalAge}_{isct})$ denotes age-at-arrival bin indicators. We estimate the second stage by replacing R_{isct} in equation (2) with its fitted value from equation (3).

A further advantage of 2SLS in this setting is that it mitigates attenuation from classical measurement error in test-based proxies for latent language proficiency. As usual, the estimates are local (LATE) to first-generation immigrant students whose reading proficiency responds to Z_{isct}.

Linguistic distance is derived from the linguistic proximity index developed by Melitz and Toubal (2014), which quantifies structural similarities between languages based on their genealogical and lexical characteristics; we compute it as one minus the proximity score.[3] The arrival age is directly obtained from the PISA datasets. This interaction term captures variation in language acquisition difficulty that arises from both linguistic and developmental factors.

[3] Appendix Table B1 summarizes the underlying CEPII language measures (COL, CNL, LP1/LP2) and show how they are combined to construct the composite index used to construct linguistic distance.

The validity of our instrument requires that the interaction of arrival age and linguistic distance affects integration and bullying outcomes only through its effect on host-country language proficiency (reading literacy). Several concerns merit discussion. First, one might worry that earlier arrival directly improves integration by providing more time to form friendships and learn social norms, independent of language. We address this by controlling flexibly for arrival age through arrival-age-bin indicators, so that identification does not come from the level effect of arriving earlier. Instead, our identification relies on whether the arrival-age gradient differs systematically with linguistic distance. Second, linguistic distance might correlate with cultural distance, which could directly affect integration through channels unrelated to language, for instance, differences in norms around peer interaction or parental involvement in schooling. We address this by controlling for cultural distance and in robustness checks including origin fixed effects and origin-destination pair fixed effects, which absorb time-invariant origin characteristics and bilateral factors common to a given origin–destination pairing. Intuitively, we compare students with similar arrival ages and from similar backgrounds but whose languages are differentially distant from the host language. Assuming linguistic distance is conditionally orthogonal to unobserved determinants of integration, the instrument isolates the component of language proficiency driven by differential learning frictions due to linguistic distance. Consistent with this interpretation, Appendix Table B2 shows that the instrument is not systematically related to predetermined characteristics once the baseline controls and fixed effects are included. We further probe sensitivity to richer fixed-effect structures and additional controls that absorb origin-related and destination-related confounding. Figure 1 illustrates the corresponding first-stage relationship between reading proficiency and the interaction of arrival age and linguistic distance.[4]

Estimation and Inference with Plausible Values Because PISA reports reading proficiency using multiple plausible values rather than a single test score, each specification is estimated separately for the ten plausible values (see Appendix A.1 for details). Coefficient estimates and standard errors are combined using Rubin's multiple-imputation rules, and standard errors are clustered at the school level.[5]

4 Results

4.1 Integration and Bullying

Baseline results Table 2 reports our main OLS and 2SLS estimates of the effects of reading proficiency on immigrant students' integration and bullying in Columns (1) and (2). The first-stage coefficient on the interaction between arrival age and linguistic distance is -0.043 for integration and -0.042 for bullying, with Kleibergen–Paap F-statistics of 69.7 and 60.2, respectively, indicating a strong first stage.

[4] Appendix Figure B1 complements Figure 1 by showing the corresponding partial relationships between reading proficiency and each component of the instrument separately after residualizing with respect to year and school fixed effects.

[5] Our conclusions are unchanged under alternative clustering levels (e.g., school×year, origin, and origin×destination).

The 2SLS estimates suggest that a one standard deviation increase in reading proficiency raises integration by 0.56 standard deviations, significant at the 1% level. By contrast, the corresponding OLS estimate is much smaller (0.065), consistent with attenuation due to measurement error in test scores and/or omitted variable bias in OLS, as students with lower latent language proficiency may also differ along unobserved dimensions relevant for integration and bullying. Appendix Table C1 reports reduced-form effects of the instrument on integration and bullying, with signs and magnitudes consistent with the 2SLS estimates.

Results for bullying follow a similar pattern. While OLS suggests a modest reduction of 0.18 standard deviations, the 2SLS estimate implies a substantially larger decline of 0.59 standard deviations, significant at the 1% level. As with integration, these 2SLS estimates should be interpreted as local average treatment effects for first-generation immigrant students whose reading proficiency responds to the interaction between arrival age and linguistic distance.[6]

To benchmark magnitudes, we scale the 2SLS estimates by observed native/immigrant gaps (Table 1). The native/immigrant gap in reading is 0.214 standard deviations, while the corresponding gaps are 0.122 standard deviations for integration and 0.057 standard deviations for bullying. Our estimates imply that closing 10% of the reading gap would close about 10% of the integration gap ($0.558 \times 0.10 \times 0.214/0.122 \approx 0.10$) and about 22% of the bullying gap ($0.589 \times 0.10 \times 0.214/0.057 \approx 0.22$). Note that 2SLS identifies a local average treatment effect for students whose reading proficiency responds to the arrival-age $\times$ linguistic-distance instrument. If these "compliers" are disproportionately those facing the largest language-acquisition frictions, marginal returns to improved proficiency may exceed average returns. For these reasons, we treat the gap-scaled calculations as a descriptive aid for interpretation rather than a literal forecast of large-scale policy interventions.

Alternative Fixed Effects and Cultural Distance Control A key concern is that the arrival-age $\times$ linguistic-distance interaction may proxy for broader origin–destination differences, such as cultural proximity, bilateral ties, or destination-specific time shocks, that may affect our outcomes of interest through channels other than language proficiency. The remaining columns of Table 2 examine whether the baseline estimates hinge on the fixed-effects structure or on cross-country cultural heterogeneity. Adding cultural distance in Column (3) leaves the second-stage coefficients almost unchanged for both outcomes, reducing concerns that unobserved cultural proximity between origin and destination countries is confounding the estimated effect of reading proficiency. Columns (4) and (5) strengthen identification by adding origin-country fixed effects and origin–destination pair fixed effects, respectively, thereby absorbing time-invariant origin-specific factors and origin–destination pair characteristics. The point estimates remain stable across these increasingly demanding specifications. Finally, Column (6) introduces destination country$\times$year fixed effects, which flexibly absorb any country-specific time shocks in the destination. The estimated effects remain close to the baseline for both integration and bullying. Across all specifications, the first-stage coefficients are stable and Kleibergen–Paap F-statistics remain comfortably above conventional thresholds, indicating that instrument relevance is not sensitive to the fixed-effects structure.

[6]For completeness, Appendix Table C2 reports the coefficients on the full set of controls in the baseline 2SLS specification.

Table 2: **The Effect of Language Proficiency on First-Generation Immigrant Students**

Model	Baseline OLS (1)	Baseline 2SLS (2)	Add Cultural Distance (3)	Origin FE (4)	Country–Origin Pair FE (5)	Country×Year FE (6)
Panel A: Dependent Variable – Integration						
First Stage						
Arrival age × Linguistic distance		-0.043***	-0.044***	-0.042***	-0.042***	-0.042***
		(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
KP F-statistic		69.67	71.02	65.23	66.01	65.14
Second Stage						
Reading score	0.065***	0.558***	0.538***	0.507***	0.505***	0.603***
	(0.012)	(0.153)	(0.151)	(0.159)	(0.158)	(0.161)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Origin FE	No	No	No	Yes	No	No
Country-Origin FE	No	No	No	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	No
Country×Year FE	No	No	No	No	No	Yes
N	14,118	14,118	14,118	14,118	14,118	14,118
Panel B: Dependent Variable – Bullying						
First Stage						
Arrival age × Linguistic distance		-0.042***	-0.042***	-0.041***	-0.041***	-0.041***
		(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
KP F-statistic		60.23	61.00	57.16	57.42	56.64
Second Stage						
Reading score	-0.182***	-0.589***	-0.573***	-0.580***	-0.570***	-0.608***
	(0.015)	(0.181)	(0.179)	(0.191)	(0.191)	(0.188)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Origin FE	No	No	No	Yes	No	No
Country-Origin FE	No	No	No	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	No
Country×Year FE	No	No	No	No	No	Yes
N	13,448	13,448	13,448	13,448	13,448	13,448

Notes: Integration (Panel A) is a standardized index of school integration (higher values indicate stronger integration), and bullying (Panel B) is a standardized index of bullying exposure (higher values indicate more bullying). Column (1) reports OLS estimates; Columns (2)–(6) report 2SLS estimates instrumenting reading proficiency with the interaction between age at arrival and linguistic distance. Column (3) additionally controls for cultural distance, while Columns (4)–(6) add origin fixed effects, destination–origin pair fixed effects, and country, year, and country×year fixed effects, respectively. All specifications control for age-at-arrival categories and continuous linguistic distance, include individual- and school-level controls, and incorporate school and year fixed effects as indicated. Standard errors are clustered at the school level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Source: PISA 2015, 2018, and 2022; CEPII; WVS.

Alternative Measure of Language Proficiency Appendix Table C3 reports robustness results using an alternative proxy for host-country language proficiency based on whether students primarily speak the host-country language at home, as reported in PISA. This measure captures a complementary dimension of everyday language exposure, albeit in a coarser form than standardized reading test scores.

Re-estimating both the OLS and IV specifications using this proxy yields results that are qualitatively consistent with the baseline findings: speaking the host-country language at home is positively associated with school integration and negatively associated with exposure to bullying, with IV estimates larger in magnitude than their OLS counterparts. Overall, these results support the interpretation that the main patterns are not specific to the reading-score measure.

Impacts on individual components Moving beyond the aggregate school integration and bullying indices, we examine the effects on the individual components underlying each index. The item-level estimates in Table 3 and Figure 2 show that for integration higher reading proficiency increases the likelihood of feeling included, welcomed and connected at school, and makes it easier to make friends. For bullying, better language skills reduce reports of being isolated, threatened, and having property destroyed, as well as teasing and fighting, with weaker and less precisely estimated effects for rumor-spreading. These patterns suggest that the main index results are broad-based rather than being driven by a single idiosyncratic item.

Table 3: **2SLS Estimates by Components of Outcome Variables**

Panel A: Integration						
	(1) Included	(2) Easy to make friends	(3) Strong belonging	(4) Comfortable	(5) Welcomed	(6) Connected
Reading score	0.553***	0.279**	0.128	0.169	0.365***	0.338***
	(0.141)	(0.124)	(0.113)	(0.116)	(0.104)	(0.127)
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
KP F-statistic	64.99	64.51	67.71	72.07	68.67	68.40
N	12,966	13,050	13,003	12,879	12,966	12,991
Panel B: Bullying						
	Isolated	Teased	Threatened	Property destroyed	Fight	Rumors
Reading score	-0.368***	-0.290**	-0.338***	-0.432***	-0.293***	-0.174
	(0.129)	(0.138)	(0.105)	(0.111)	(0.103)	(0.115)
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
KP F-statistic	57.71	58.85	60.05	58.57	58.90	59.43
N	13,338	13,337	13,313	13,322	13,346	13,329

Note: All items are coded so that higher values indicate stronger social integration (Panel A) or greater exposure to bullying (Panel B). Reading proficiency is instrumented with the interaction between age at arrival and linguistic distance between origin and destination languages. All specifications control for age-at-arrival categories and continuous linguistic distance, include student- and school-level controls, and incorporate school and year fixed effects. Standard errors (in parentheses) are clustered at the school level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Source: PISA 2015, 2018 and 2022; CEPII.

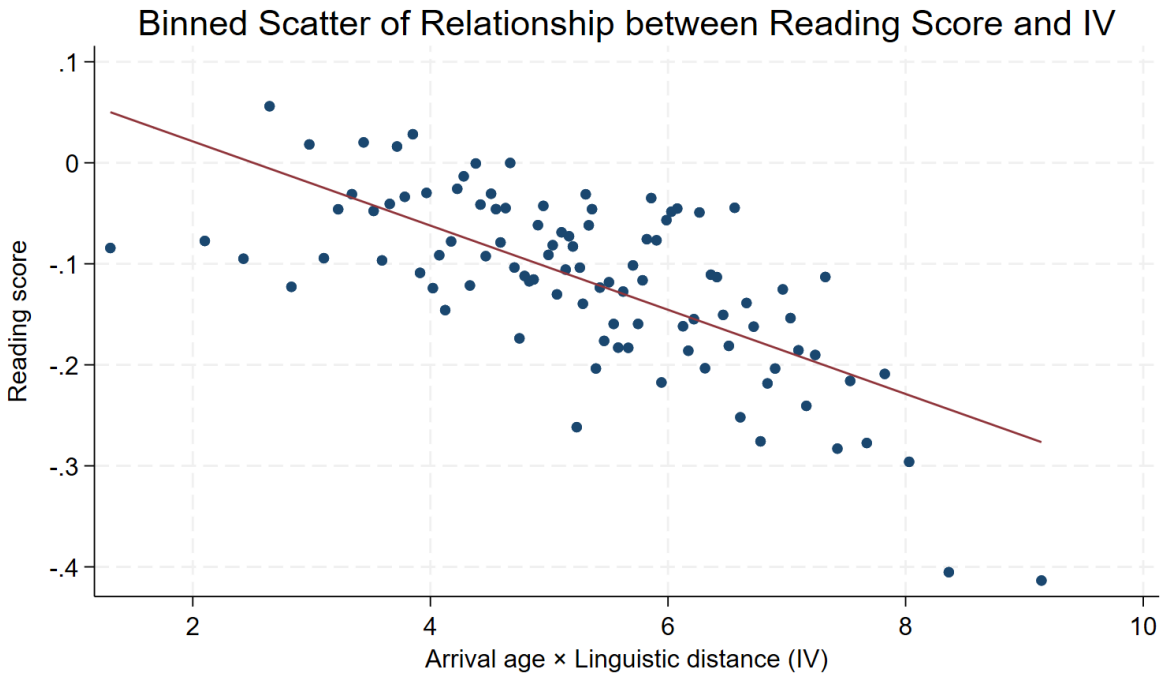
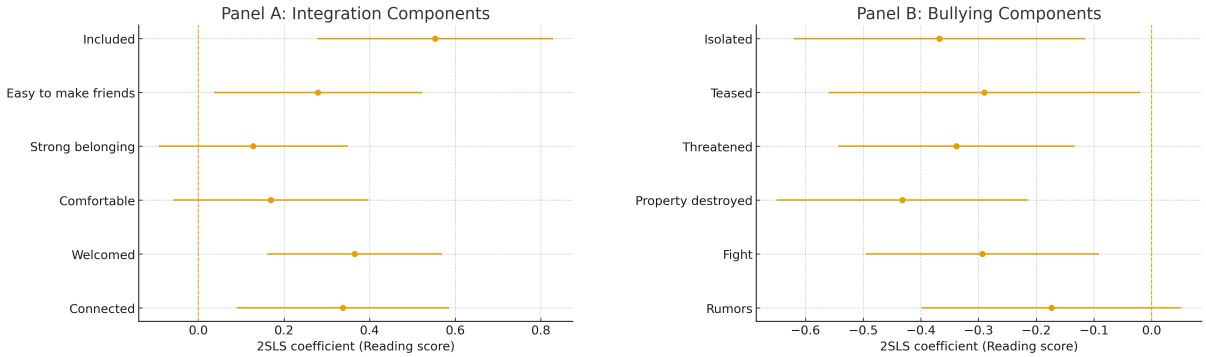


Figure 1: **First-stage relationship between Z_{isct} = ArrivalAge$_{isct}$ × LingDist$_{isct}$ and reading proficiency**
Source: PISA 2015, 2018 and 2022, CEPII.



(a) **Integration components** (b) **Bullying components**

Note: Panel (a) reports the 2SLS estimates of the effect of reading proficiency on six components of social integration. Panel (b) presents the corresponding effects on different forms of bullying victimization. Standard errors are clustered at the school level.
Source: PISA 2015, 2018, 2022; CEPII.

Figure 2: **Effects of Language Proficiency on Components of Integration and Bullying**

4.2 Heterogeneity Analysis

Table 4 examines whether the effects of language proficiency vary by school immigrant share, gender, and age at arrival. For each specification, we report both the main effect and the corresponding interaction term, allowing us to assess heterogeneity in effects across groups.

We find little evidence of heterogeneity by immigrant concentration for integration. For bullying, however, the interaction between reading proficiency and the school-level immigrant share is positive and statistically significant at the 10% level, indicating that the protective effect of language proficiency against bullying is somewhat weaker in schools with a higher share of immigrant students. The magnitude of this heterogeneity is modest. One possible interpretation is that in schools with a large immigrant presence, linguistic barriers may be less stigmatizing because they are more common, thereby attenuating the marginal benefit of individual language proficiency.

For gender, the negative interaction with a female indicator for integration (-0.236) suggests the integration benefits of language proficiency are somewhat smaller for girls. The positive coefficient on the gender interaction for bullying (0.435) indicates language proficiency is more protective against bullying for boys. The gender difference is quantitatively large: the implied effect of a one-standard-deviation increase in reading proficiency is approximately 0.78 SD for boys but only 0.34 SD for girls. This pattern may reflect differences in the nature of bullying experienced by boys and girls, as boys are more likely to face physical forms of bullying that may be more sensitive to peer-group integration and communication skills. Appendix Table A5 shows that, within the immigrant sample, boys have substantially lower reading proficiency and report markedly higher bullying exposure than girls. This baseline pattern helps interpret why the marginal protective effect of language proficiency against bullying is stronger for boys.

Interactions with a late-arrival indicator (arrival after age 8) are less precisely estimated, pointing to broadly similar language effects across arrival-age groups. However, a weak first stage (F-statistics of 2.0 and 2.5), renders these estimates unreliable.

4.3 Secondary outcomes

We also explore the effects of language proficiency on other academic and school behaviors. Results reported in Table 5 show that language proficiency has important consequences for academic performance and grade progression. A one-standard-deviation increase in reading proficiency raises mathematics scores by about 0.42 standard deviations and science scores by about 0.82 standard deviations. It also reduces the probability of having repeated a grade by around 12 percentage points, relative to a mean repetition rate of 0.165 among first-generation immigrants. In contrast, the estimated effect on absenteeism is negative but imprecisely estimated.

Note that Isphording et al. (2016), using the same instrumental-variable strategy with earlier PISA waves (2003–2012), report a 2SLS coefficient of 0.57 standard deviations for the effect of reading on mathematics. Our somewhat smaller estimate likely reflects differences in sample composition and migration contexts across waves and destination countries, but the broad consistency supports the comparability of our approach.

Table 4: **Heterogeneity Analysis**

Dependent variable	Integration			Bullying		
	(1)	(2)	(3)	(4)	(5)	(6)
Reading score	0.480***	0.672***	-0.230	-0.883***	-0.777***	0.929
	(0.180)	(0.171)	(0.715)	(0.233)	(0.198)	(0.898)
Immigration share × Reading	0.147			0.568*		
	(0.238)			(0.297)		
Female × Reading		-0.236**			0.435***	
		(0.096)			(0.110)	
Arrival age over 8 × Reading			0.472			-0.922
			(0.475)			(0.604)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
KP F-statistic	34.27	34.62	2.01	29.16	29.99	2.45
N	14,118	14,118	14,118	13,448	13,448	13,448

Notes: This table reports 2SLS estimates of the effect of reading proficiency on students' integration and bullying outcomes, allowing for heterogeneity by school immigrant share, gender, and an indicator for arrival age above eight. Integration is a standardized index of school belonging, and bullying a standardized index of bullying victimization. Reading proficiency is instrumented with the interaction between age at arrival and linguistic distance between origin and destination languages; interaction terms are instrumented by interacting the corresponding heterogeneity variables with the instrument. All specifications control for arrival-age categories and continuous linguistic distance, include student- and school-level controls, and incorporate school and year fixed effects. Standard errors (in parentheses) are clustered at the school level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Source: PISA 2015, 2018, 2022; CEPII.

Table 5: **2SLS Results: Effect of Reading Proficiency on Academic and Behavioural Outcomes**

	Math Score	Science Score	Repeat Grade	Absenteeism
	(1)	(2)	(3)	(4)
Second Stage				
Reading score	0.419***	0.818***	-0.115**	-0.125
	(0.103)	(0.093)	(0.047)	(0.079)
First Stage				
Arrival age × linguistic distance	-0.043***	-0.043***	-0.043***	-0.042***
	(0.006)	(0.006)	(0.006)	(0.008)
KP F-statistic	69.67	69.74	68.26	37.41
Controls	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	14,118	14,118	14,032	9,172

Notes: Reading proficiency is instrumented with the interaction between continuous age at arrival and linguistic distance. Mathematics and science are standardized PISA achievement scores; grade repetition and absenteeism are binary indicators for having ever repeated a grade or skipped any class. All specifications control for age-at-arrival categories and continuous linguistic distance, include student- and school-level controls, and incorporate school and year fixed effects. Standard errors (in parentheses) are clustered at the school level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Source: PISA 2015, 2018, 2022; CEPII.

4.4 Spillovers to Native Students

We further examine whether immigrant peers' language proficiency is associated with native students' academic and psychosocial outcomes. Exploiting variation across schools and PISA waves, we construct two school-level peer measures for each school-wave cell: (i) the average reading proficiency of immigrant students, computed as the sampling-weight–adjusted mean of the ten PISA plausible values ($ImmRead_{st}$); and (ii) the share of immigrant students in the school ($ShareMig_{st}$). We also compute the school-wave average age at arrival of immigrant students, which we use as an instrument for $ImmRead_{st}$. We then estimate the following specification:

$$\mathrm{Y}_{ist} = \beta_0 + \beta_1 \cdot \mathrm{ImmRead}_{st} + \beta_2 \cdot \mathrm{ShareMig}_{st} + \mathbf{X}'_{ist}\boldsymbol{\gamma} + \varphi_s + \lambda_{c*t} + \varepsilon_{ist}, \tag{4}$$

where Y_{ist} denotes the psychosocial outcomes (integration and bullying), academic performance (reading, mathematics, and science) and educational and occupational expectations of native student i in school s and wave t. The key variable $\mathrm{ImmRead}_{st}$ captures the average reading proficiency of immigrant students in the same school-year cell, while $\mathrm{ShareMig}_{st}$ controls for the overall share of immigrant peers. The vector $\mathbf{X}_{ist}$ includes individual, parental, and school covariates. School fixed effects and country×year fixed effects absorb time-invariant school characteristics and time-varying country-level shocks common to all schools within a country.

We instrument school-average immigrant reading with school-average immigrant arrival age. Identification comes from within-school variation across PISA waves in the arrival-age composition of immigrant students, which predicts changes in immigrant students' average reading proficiency. However, the exclusion restriction is less compelling at the school level than in the individual-level design: immigrant arrival patterns may co-move with other changes in time-varying school composition or local shocks that directly affect native students. We therefore view these results as suggestive evidence consistent with peer spillovers, but cannot rule out that these effects operate partly through other channels.

Table 6 reports the OLS and 2SLS results. The 2SLS results suggest that native students in schools where immigrant peers have higher reading proficiency report better integration and lower bullying exposure, and achieve higher reading, mathematics, and science scores. A one-standard-deviation increase in immigrant peers' average reading proficiency is associated with roughly 0.13–0.18 standard-deviation gains in natives' reading, mathematics, and science test scores and modest improvements in their educational and occupational expectations. These patterns suggest that language proficiency among immigrant students is not only privately beneficial but also associated with a more supportive and academically productive school environment for native peers.

4.5 Mechanisms

The results so far show that host-country language proficiency has large effects on immigrant students' school integration and exposure to bullying. In this subsection, we ask why better language skills translate into better inclusion. We focus on three related channels: improved relationships with peers, improved relationships with teachers, and a broader bundle of socio-emotional skills and expectations that support participation in school life. While our empirical

Table 6: **OLS and 2SLS Results: Spillovers of Immigrant Students' Language Proficiency to Native Students**

	Integration	Bullying	Reading	Math	Science	Expected education years	Expected job level
Panel A: OLS							
School-average immigrants' reading score	0.050***	-0.049***	0.211***	0.191***	0.198***	0.221***	0.129***
	(0.005)	(0.005)	(0.007)	(0.008)	(0.007)	(0.015)	(0.007)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.062	0.067	0.341	0.353	0.335	0.300	0.254
N	251,553	237,139	259,699	259,699	259,699	237,669	189,638
Panel B: 2SLS							
Second stage:							
School-average immigrants' reading score	0.106***	-0.106***	0.129***	0.175***	0.135***	0.315***	0.140***
	(0.025)	(0.025)	(0.032)	(0.036)	(0.033)	(0.069)	(0.034)
First stage:							
School-average immigrant arrival age	-0.043***	-0.043***	-0.043***	-0.043***	-0.043***	-0.043***	-0.041***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
KP F-statistic	159.94	152.61	160.52	160.52	160.52	150.84	143.71
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	251,553	237,139	259,699	259,699	259,699	237,669	189,638

Notes: Integration and bullying are standardized indices; reading, mathematics, and science are standardized PISA test scores. School-level average immigrant reading proficiency is instrumented by the school–year average immigrant age at arrival. All specifications include school and country×year fixed effects. Standard errors are clustered at the school level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Source: PISA 2015, 2018, and 2022.

design does not allow us to cleanly separate these channels, several patterns in the data are informative and suggestive. Appendix Table A6 reports summary statistics and survey-wave availability for each mechanism variable.

We begin by examining outcomes related to expectations, study time, teacher relationships, and socio-emotional skills. Table 7 reports the corresponding OLS results. Panel A shows that reading proficiency is strongly and positively associated with expected years of education, expected occupational status at age 30, and expected occupational status relative to parents.

Panel B documents negative associations between reading proficiency and out-of-school study time and homework/study time at home. This pattern is consistent with a productivity channel: students with better language skills can complete a given amount of work in less time, rather than simply reducing effort.

Panel C focuses on teachers and the school climate. We see that higher reading proficiency is associated with substantially lower reported unfair treatment by teachers and higher student–teacher relationship quality. Finally, Panel D reports associations with socio-emotional skills, global competences, and psychological traits. Students with higher reading proficiency score higher on resilience, attitudes toward immigrants, cognitive flexibility, respect for other cultures, intercultural communication awareness, global-mindedness, empathy, assertiveness, emotional control, and growth mindset, and report lower mathematics anxiety.

These findings suggest that the benefits of language proficiency extend far beyond traditional human capital measures. Consistent with Heckman and Kautz (2012), we find that language does not just improve test scores but improves the non-cognitive profile, which in turn can act as a protective factor against peer victimization. By improving a student's ability to engage with the school environment, language proficiency may support the 'joint production' of cognitive and socio-emotional skills.

We interpret these OLS results as suggestive evidence on mechanisms rather than as identified channels, as they may reflect both causal effects of language proficiency and selection on unobserved traits. With this limitation in mind, the patterns in Table 7 are consistent with the channels outlined earlier: host-country language proficiency appears closely linked to higher aspirations, greater study productivity, more positive teacher relationships, and stronger socio-emotional skills, providing a plausible explanation for the large causal effects on integration and bullying documented in our 2SLS results.

Table 7: **OLS Results: Immigrant Students' Language Proficiency Effects across Educational, Labor-market and Non-cognitive Outcomes**

Outcome	Coef. of reading	SE	N	Wave
Panel A: Educational and occupational expectations				
Expected years of education	0.432***	0.028	12,697	2015, 2018, 2022
Expected occupational status	4.882***	0.255	10,156	2015, 2018, 2022
Expected occupational status relative to parents	2.013***	0.398	10,156	2015, 2018, 2022
Panel B: Study attitudes and time				
Out-of-school study time (hours/week)	-3.156***	0.395	2,867	2015
Homework/study time at home	-0.375***	0.073	4,305	2015
Panel C: Teacher relationships and school climate				
Teacher's unfairness treatment	-0.994***	0.114	3,514	2015
Student–teacher relationship quality	0.179***	0.031	3,208	2022
Panel D: Socio-emotional skills, global competences, and psychological traits				
Resilience	0.085***	0.022	4,753	2018
Attitudes toward immigrants	0.247***	0.021	3,425	2018
Cognitive flexibility	0.061**	0.025	3,591	2018
Respect for other cultures	0.221***	0.021	3,496	2018
Intercultural communication awareness	0.224***	0.023	3,562	2018
Global-mindedness	0.123***	0.025	3,462	2018
Empathy	0.089***	0.026	4,066	2022
Assertiveness	0.064***	0.024	3,967	2022
Emotional control	0.105***	0.026	3,687	2022
Growth mindset	0.192***	0.025	3,958	2022
Mathematics anxiety	-0.181***	0.029	3,607	2022

Notes: Each coefficient reports the OLS effect of a one–standard-deviation increase in reading proficiency on the corresponding outcome. Panel A reports educational and occupational expectations (expected years of education and anticipated occupational status). Panel B examines study time and learning-related behaviors outside school. Panel C focuses on perceptions of teacher fairness and student–teacher relationships. Panel D presents socio-emotional skills, global competences, and psychological traits. Outcomes in Panels C and D are standardized within survey wave to mean zero and unit variance. Expected years of education (Panel A) and out-of-school study time (Panel B) are measured in natural units (years and hours per week), while other outcomes in Panels A and B are reported in their original scales. All specifications control for age-at-arrival categories and continuous linguistic distance, include student- and school-level controls, and incorporate school and year fixed effects. Standard errors are clustered at the school level. *, **, and *** denote significance at the 10%, 5%, and 1% levels. "Wave" denotes the PISA survey year(s) in which each outcome is observed.

Source: PISA 2015, 2018 and 2022

4.6 Second-Generation Immigrant Students

Table 8 extends the analysis to second-generation immigrant students, for whom language proficiency remains a salient dimension of integration despite being born in the host country. Ex-

amining this group is important to assess whether the relationship between language skills and social inclusion extends beyond recent arrivals and reflects broader integration challenges. For this subgroup our IV strategy is not applicable, because the identifying variation in age-at-arrival does not apply, thus we present OLS estimates. For these students, we find consistent patterns to those of first generation student. Specifically, a one-standard-deviation increase in reading proficiency is associated with a 0.05-standard-deviation increase in integration and a 0.15-standard-deviation reduction in bullying, both precisely estimated. We interpret these as evidence that language proficiency is also an important correlate of social inclusion for children of immigrants who were born in the host country, even though we do not attempt to identify causal effects in this group.

Table 8: **OLS Results: The Effect of Language Proficiency on School Integration among Second-Generation Immigrant Students**

Dependent Variable	Integration	Bullying
	(1)	(2)
Reading score	0.045***	-0.150***
	(0.005)	(0.005)
Controls	Yes	Yes
School FE	Yes	Yes
Year FE	Yes	Yes
R^2	0.125	0.136
N	97,882	93,057

Notes: This table reports OLS estimates of the effect of language proficiency on school integration and bullying among second-generation immigrant students. All specifications include student- and school-level covariates as well as school and year fixed effects. Standard errors (in parentheses) are clustered at the school level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Source: PISA 2015, 2018, and 2022.

5 Robustness

We conduct a number of robustness checks along three dimensions: (i) sensitivity to how outcomes are constructed, (ii) sensitivity to sample definition and missing-data handling, and (iii) sensitivity to inference choices and to omitted-variable concerns. Table 9 assesses the sensitivity of the main results to a wide range of alternative specifications, each modifying a key aspect of the empirical setup, including outcome construction, sample composition, treatment of missing data, and the handling of the survey design, to gauge the stability of the estimates in both Panel A (integration) and Panel B (bullying).

Outcome construction Column (2) replaces the additive indices of integration and bullying with PCA-based measures constructed from the underlying items. The resulting 2SLS coefficients remain very close to the baseline in both panels, with comparable levels of precision, indicating that the main findings are not driven by the particular index-construction rule.

Sample definition Column (3) expands the sample to include non-OECD destination countries, substantially increasing institutional and socioeconomic heterogeneity. Despite this broader context, the estimated effects of reading proficiency remain highly similar to those in the baseline specification, suggesting that the relationship between language proficiency and psychosocial outcomes is not specific to OECD destinations.

Weights Column (4) re-estimates the model using PISA's student sampling weights. While standard errors increase modestly, as expected when accounting for unequal selection probabilities, the point estimates remain largely unchanged in both panels. This indicates that the baseline results are not sensitive to weighting.

Missing data Column (5) restricts the analysis to the complete-case sample by dropping all observations with imputed values. Although this substantially reduces the effective sample size and leads to slightly noisier estimates, the coefficients on reading proficiency remain close to the baseline in both magnitude and significance. This pattern suggests that multiple imputation is not driving the main results.

Covariates Columns (6) and (7) provide complementary stress tests: dropping the covariate set entirely versus augmenting it with richer school-level controls. Column (6) estimates the baseline model without any individual- or school-level covariates, while retaining school and year fixed effects. The resulting estimates retain the same sign in both panels, though the bullying coefficient attenuates substantially relative to the baseline, indicating that the bullying magnitude is more sensitive to conditioning on observables than the integration estimate. Column (7) augments the baseline specification with a richer set of school-level structural controls, such as school size, teacher–student ratios, class size, and teacher shortage indicators, and the ESCS index. This specification addresses potential concerns related to incomplete parental background information at the individual level and unobserved heterogeneity in school resources and learning environments. Although the inclusion of these additional controls substantially reduces the estimation sample, due to missingness in these richer school-resource measures across schools/waves, the resulting 2SLS estimates remain close in magnitude and statistical significance to the baseline.

Inference Column (8) incorporates the full PISA survey design by using Fay's Balanced Repeated Replication (BRR) replicate weights for variance estimation. As expected, BRR-based standard errors are somewhat larger than those obtained under clustered inference, but the point estimates remain close to the benchmark results in both panels.

Overall, the results in Table 9 show that the main coefficients are stable in sign and similar in magnitude across a wide range of alternative specifications, with integration estimates consistently around 0.46-0.74 and bullying estimates around -0.25 to -0.68 across the variants.

Appendix Table D1 reports country-specific estimates. Effects are broadly consistent across major receiving countries showing strong first stages and second-stage coefficients that generally match with the pooled results. However, several smaller destinations yield imprecise estimates due to limited sample sizes and weak first stages (e.g. Austria, Switzerland, and the United Kingdom, where F-statistics fall below conventional thresholds), but no single country drives the

pooled findings. We view these results primarily as a check against any single destination driving the pooled estimates, rather than as stand-alone evidence for smaller destinations.

Finally, Appendix Table D2 probes sensitivity to how we parameterize arrival age and linguistic distance (bins vs. continuous; continuous distance vs. top-quartile indicator). Integration estimates remain positive across variants, while bullying magnitudes attenuate and become less precise under tighter parameterizations that absorb more identifying variation.

Table 9: **Robustness Check: Different Specifications**

	(1) **Baseline**	(2) **PCA**	(3) **With Non-OECD**	(4) **Weighted**	(5) **Non-Imputed**	(6) **Without Controls**	(7) **Rich Controls**	(8) **BRR**
Panel A: Integration								
First Stage								
Arrival age × linguistic distance	-0.043***	-0.043***	-0.039***	-0.039***	-0.045***	-0.054***	-0.047***	-0.039***
	(0.006)	(0.006)	(0.003)	(0.008)	(0.007)	(0.003)	(0.008)	(0.004)
KP F-statistic	69.67	69.36	303.10	27.76	68.11	439.51	43.21	65.56
Second Stage								
Reading score	0.558***	0.515***	0.542***	0.594**	0.472***	0.460***	0.743***	0.593***
	(0.153)	(0.149)	(0.072)	(0.247)	(0.151)	(0.055)	(0.207)	(0.226)
Controls	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	14,118	14,058	24,748	14,118	12,513	14,118	7,866	14,118
Panel B: Bullying								
First Stage								
Arrival age × linguistic distance	-0.042***	-0.042***	-0.039***	-0.038***	-0.043***	-0.053***	-0.045***	-0.039***
	(0.006)	(0.006)	(0.003)	(0.008)	(0.007)	(0.003)	(0.008)	(0.004)
KP F-statistic	60.230	59.180	283.92	24.950	57.100	411.630	37.420	58.350
Second Stage								
Reading score	-0.589***	-0.672***	-0.526***	-0.551*	-0.465***	-0.246***	-0.681***	-0.577**
	(0.181)	(0.192)	(0.078)	(0.288)	(0.178)	(0.062)	(0.257)	(0.230)
Controls	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	13,448	13,417	23,529	13,448	11,958	13,448	7,495	13,448

Notes: Panel A reports results for the integration index, and Panel B for the bullying index. Column (1) reports the baseline specification. Column (2) replaces the outcome with principal-component-based indices constructed from the underlying psychosocial items. Column (3) expands the sample to include non-OECD destination countries. Column (4) re-estimates the model using PISA student sampling weights. Column (5) restricts the sample to non-imputed observations. Column (6) excludes individual- and school-level covariates while retaining school and year fixed effects. Column (7) adds an extended set of school-level structural controls and the PISA index of economic, social, and cultural status (ESCS). Column (8) uses Fay's balanced repeated replication (BRR) replicate weights for variance estimation. In all specifications, reading proficiency is instrumented with the interaction between age at arrival and linguistic distance. Standard errors (in parentheses) are clustered at the school level, except in Column (8), which reports BRR-based standard errors using Fay's replicate weights. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Source: PISA 2015, 2018, and 2022; CEPII.

6 Conclusion

This paper speaks to persistent disparities in host country language proficiency among immigrant students and the role these skills play in shaping both learning and social participation in school. Although immigrant-background children represent a growing share of students in OECD countries, substantial variation exists in their exposure to the host-country language, with only around half speaking it at home, see (OECD, 2025). At the same time, immigrant children typically underperform native peers in reading and other academic outcomes. This raises important questions about how language proficiency influences not only academic achievement but also the social and emotional dimensions of schooling. Understanding its causal impact on school integration is therefore critical for explaining persistent inequalities in human capital formation

and for informing policies that support the successful inclusion of immigrant students.

We provide new evidence that host-country language proficiency plays a central role in shaping immigrant students' academic performance, social integration, and emotional well-being at school. Using harmonized PISA data from 16 OECD destination countries and an instrumental-variable strategy based on age at arrival and linguistic distance between immigrants' mother tongue and the host-country language, we estimate large causal effects of language proficiency on school integration and exposure to bullying among first-generation immigrant students. A one-standard-deviation increase in reading proficiency raises a standardized integration index by about 0.56 standard deviations and reduces a standardized bullying index by about 0.59 standard deviations. These effects are sizable and substantially larger than corresponding OLS estimates, consistent with attenuation and selection bias in naive specifications.

The impacts of language proficiency differ across dimensions of integration. The protective effect against bullying is particularly strong for boys, whereas improvements in school integration are more evenly distributed across genders. Beyond social inclusion, higher language proficiency also improves academic outcomes, increasing achievement in mathematics and science and reducing the probability of grade repetition. Together, these results highlight that language proficiency is not merely an academic input but also matters for social inclusion and well-being at school.

The benefits of immigrant students' language proficiency also extend beyond immigrants themselves. We document evidence consistent with positive peer spillovers for native students: natives in schools where immigrant peers have higher average reading proficiency report better integration and lower exposure to bullying, and better educational performance. These findings suggest that language proficiency among immigrant students may contribute to a more supportive, inclusive, and academically productive school environment, generating potential positive externalities for the wider student body. For second-generation immigrant students, OLS estimates reveal qualitatively similar patterns, with higher reading proficiency associated with better integration and lower bullying, suggesting that language skills remain closely linked to social inclusion even among children born in the host country. Exploring potential mechanisms, we show that higher reading proficiency is associated with reduced communication frictions, stronger engagement with teachers, and enhanced socio-emotional capacities that facilitate participation in school life.

Overall, the results demonstrate that the returns to host-country language proficiency extend well beyond test scores. Language skills enable immigrant students to participate more fully in the social fabric of schools, reduce vulnerability to negative peer interactions, and strengthen socio-emotional well-being—factors that are themselves predictive of long-run educational and labor-market outcomes. From a policy perspective, the findings imply that early and effective investments in language support for immigrant children can yield high returns, not only by improving academic achievement but also by fostering inclusive school environments. As immigrant-background students account for a growing share of school populations across OECD countries, language acquisition policies represent a powerful lever for promoting integration, reducing inequality, and supporting both cognitive and psychosocial development.

References

Alan, S., Baysan, C., Gumren, M. and Kubilay, E. (2021), 'Building social cohesion in ethnically mixed schools: An intervention on perspective taking', *Quarterly Journal of Economics* **136**, 2147–2194.

Bleakley, H. and Chin, A. (2004), 'Language skills and earnings: Evidence from childhood immigrants', *Review of Economics and Statistics* **86**(2), 481–496.

Bleakley, H. and Chin, A. (2010), 'Age at arrival, english proficiency, and social assimilation among us immigrants', *American Economic Journal: Applied Economics* **2**(1), 165–192.

Boucher, V., Tumen, S., Vlassopoulos, M., Wahba, J. and Zenou, Y. (2021), Ethnic mixing in early childhood: Evidence from a randomized field experiment and a structural model. CEPR Discussion Paper #15528.

Brown, S. and Taylor, K. (2008), 'Bullying, education and earnings: evidence from the national child development study', *Economics of Education Review* **27**(4), 387–401.

Cavallo, M. and Russo, G. (2025), 'Lost in translation: Reading performance and math performance of second-generation immigrant children in italy', *Journal of Human Capital* **19**(1), 80–114.

Chiswick, B. R. and Miller, P. W. (1995), 'The endogeneity between language and earnings: International analyses', *Journal of Labor Economics* **13**(2), 246–288.

Clarke, A. and Isphording, I. E. (2017), 'Language barriers and immigrant health', *Health Economics* **26**(6), 765–778.

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. and van Soest, A. (2001), 'Language fluency and earnings: Estimation with misclassified language indicators', *Review of Economics and Statistics* **83**, 663–674.

Entorf, H. (2015), 'Migrants and educational achievement gaps', *IZA World of Labor* .

Eriksen, T. L. M., Nielsen, H. S. and Simonsen, M. (2014), 'Bullying in elementary school', *Journal of Human Resources* **49**(4), 839–871.

Fearon, J. D. and Laitin, D. D. (2003), 'Ethnicity, insurgency, and civil war', *American Political Science Review* **97**(1), 75–90.

Geay, C., McNally, S. and Telhaj, S. (2013), 'Non-native speakers of english in the classroom: What are the effects on pupil performance?', *The Economic Journal* **123**(570), F281–F307.

Heckman, J. J. and Kautz, T. (2012), 'Hard evidence on soft skills', *Labour Economics* **19**(4), 451–464.

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.

Maynard, B. R., Vaughn, M. G., Salas-Wright, C. P. and Vaughn, S. (2016), 'Bullying victimization among school-aged immigrant youth in the united states', *Journal of Adolescent Health* **58**(3), 337–344.

Melitz, J. and Toubal, F. (2014), 'Native language, spoken language, translation and trade', *Journal of International Economics* **93**(2), 351–363.

OECD (2018), *The resilience of students with an immigrant background: Factors that shape well-being*, OECD Publishing, Paris.

OECD (2025), *Migration Data Brief @OECD no 15*, OECD Publishing, Paris.

Ortega, A. and Ludwig, T. (2023), 'Immigrant english proficiency, children's educational performance, and parental involvement', *Review of Economics of the Household* **21**(2), 693–719.

Pottie, K, D. G. G. K. P. K. and Hassan, G. (2015), 'Do first generation immigrant adolescents face higher rates of bullying, violence and suicidal behaviours than do third generation and native born?', *Journal of Immigrant and Minority Health* **17**(3), 1557–1566.

Sarzosa, M. (2024), 'Victimization and skill accumulation: the case of school bullying', *Journal of Human Resources* **59**(1), 242–279.

Schilling, P. and Höckel, L. S. (2025), 'Starting off on the right foot–language learning classes and the educational success of refugee children', *The Economic Journal* p. ueaf087.

Schmid, L. (2023), 'The impact of host language proficiency on migrants' employment outcomes', *American Economic Review: Insights* **5**(4), 477–493.

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

Swadesh, M. (1952), 'Lexico-statistic dating of prehistoric ethnic contacts: with special reference to north american indians and eskimos', *Proceedings of the American philosophical society* **96**(4), 452–463.

Tumen, S., Vlassopoulos, M. and Wahba, J. (2025), The power of language: Educational and mental health impacts of language training for refugee children. IZA Discussion Paper #17593.

Appendix

A Data and Variable Definitions

A.1 Measurement of Language Proficiency

PISA does not report a single "true" proficiency score for each student. Instead, proficiency is modelled as a latent trait within a large-scale item response theory (IRT) framework. Based on students' item responses and background characteristics, PISA first derives the posterior distribution of each student's latent reading ability. To appropriately reflect measurement uncertainty, ten *plausible values* (PVs) are randomly drawn from this posterior distribution. Each PV represents one possible realisation of the student's latent skill. Therefore, using only a single PV, or averaging the PVs prior to analysis, would underestimate measurement uncertainty and yield standard errors that are too small. Following OECD's recommended procedures, all analyses incorporate the full set of ten PVs.

Mean and Variation Descriptive statistics are calculated using the student sampling weight and the 80 replicate weights supplied by PISA. We employ Fay's Balanced Repeated Replication (BRR, Fay factor $\rho = 0.5$) to obtain unbiased population-representative estimates and valid sampling variances. Means and variances therefore reflect both the complex multistage sampling design and the stratification of PISA.

Regression Models with Plausible Values All regression models are estimated separately for each of the ten PVs. Let $\widehat{\beta}_m$ be the coefficient estimated from PV m ($m = 1, \ldots, 10$), and $\widehat{V}_m$ the associated variance computed under the main weight and BRR replicate weights. Rubin's multiple-imputation rules are then used to combine the results:

Point estimate:

$$\bar{\beta} = \frac{1}{10} \sum_{m=1}^{10} \widehat{\beta}_m. \tag{5}$$

Total variance:

$$T = \frac{1}{10} \sum_{m=1}^{10} \widehat{V}_m + \left(1 + \frac{1}{10}\right) \frac{1}{9} \sum_{m=1}^{10} (\widehat{\beta}_m - \bar{\beta})^2. \tag{6}$$

where $\widehat{V}_m$ denotes the *within-imputation variance* of the estimated coefficient in imputation m, and the second term captures the *between-imputation variance* arising from the dispersion of point estimates $\widehat{\beta}_m$ around their average $\bar{\beta}$ across imputations.

Standardisation of Reading Proficiency Because proficiency is represented by ten PVs rather than one score, standardization is conducted at the level of each plausible value. For each PV_m, we compute the weighted mean μ_m and weighted standard deviation σ_m using the student sampling weight. The standardized score for student i is

$$z_{im} = \frac{PV_{im} - \mu_m}{\sigma_m}. \tag{7}$$

Thus, each PV is standardized separately, ensuring that scaling is consistent with PISA's sampling design. The regression models are then estimated using these ten standardized PVs, and the resulting coefficients and standard errors are combined using Rubin's rules. By standardizing each PV separately and combining estimates through multiple-imputation rules, our procedure fully preserves (i) sampling uncertainty and (ii) measurement uncertainty in the latent proficiency scale.

A.2 Variable Definitions

Table A1: **Description of Variables**

Variable	Description	Note
Language Proficiency		
Reading Score	PISA reading proficiency score based on ten plausible values.	Standardized
Speak the Host Language at Home	Whether the student speaks the destination language at home.	1=Yes; 0=No
Dependent Variables		
Integration	Average of school-belonging items.	Standardized within country-by-wave
Component items:		Scale: 1=Strongly agree; 4=Strongly disagree
feel_outsider	"I feel like I am an outsider at school."	
feel_belong	"I feel like I belong at school."	
feel_lonely	"I feel lonely at school."	
feel_easy_make_friend	"I make friends easily at school."	
feel_awkward	"I feel awkward at my school."	
feel_welcome	"Other students seem to like me."	
Bullying	Average of bullying victimization items.	Standardized within country-by-wave
Component items:		Scale: 1=Never; 4=Once a week or more
bullying_isolate	"Other students left me out of things on purpose."	
bullying_tease	"Other students made fun of me."	
bullying_threat	"I was threatened by other students."	
bullying_destroy	"Other students took or destroyed my things."	
bullying_hit	"I got hit or pushed around by other students."	

Continued on next page

Variable	Description	Note
bullying_rumor	"Other students spread nasty rumours about me."	
Other Outcomes		
Math Score	PISA mathematics proficiency score based on ten plausible values.	Standardized
Science Score	PISA science proficiency score based on ten plausible values.	Standardized
Absenteeism	Students' absence rates.	Dummy
Expected Years of Education	Expected total years of schooling reported by students.	Continuous
Expected Occupational Status	Students' expected occupational status (SEI).	ISEI scale (continuous)
Expected Occupational Status Gap	Difference between expected and parental SEI.	Continuous
Out-of-School Study Time	Total hours per week spent studying outside school.	Sum index (0–70)
Homework / Study at Home	Frequency of studying or doing homework at home.	1=Never; 4=Very often
Teacher Unfairness	Index of perceived unfair treatment by teachers.	Sum index
Student–Teacher Relationship	Quality of student–teacher relationships.	Standardized
Resilience	Students' perceived resilience.	Standardized
Attitudes Towards Immigrants	Students' attitudes towards immigrants.	Standardized
Cognitive Flexibility	Adaptability and cognitive flexibility.	Standardized
Respect for Other Cultures	Respect for cultural diversity.	Standardized
Intercultural Communication Awareness	Awareness of intercultural communication.	Standardized
Global-Mindedness	Global awareness and responsibility.	Standardized
Empathy	Empathy agreement scale.	Standardized
Assertiveness	Assertiveness socio-emotional scale.	Standardized
Emotional Control	Emotional self-regulation.	Standardized
Growth Mindset	Belief in the malleability of ability.	Standardized
Mathematics Anxiety	Anxiety towards mathematics.	Standardized
Individual Characteristics		
Age	Student age in years.	

Continued on next page

Variable	Description	Note
Gender	Female indicator.	1=Female; 0=Male
Grade	International grade level (PISA derived).	
Mother/Father Education	Highest education level of parents.	0=Lower primary; 1=Primary; 2=Lower secondary; 3=Upper secondary; 4=Vocational upper secondary
Parental Occupational Status	Parents' occupational ISEI status.	Continuous (10–90)
Number of Books	Number of books at home.	1=0–10; 2=11–25; 3=26–100; 4=101–200; 5=201–500; 6=500+
Arrival Age	Age at first arrival in destination country.	
Linguistic Distance	Linguistic distance between origin and destination.	0–1 scale
Cultural Distance	Cultural distance between origin and destination.	0–1 scale
School Characteristics		
School Location	Size of community where the school is located.	1=Village; 2=Small town; 3=Town; 4=City; 5=Large city
Public School	Whether the school is publicly managed.	1=Public; 0=Private
Share of School Immigration	The share of immigrants at school.	Standardized
School-Average Immigrants' Reading Score	The weighted average reading score of immigrants at school.	Standardized
School-Average Immigrant Arrival Age	The weighted average arrival age of immigrants at school.	

Note: This table provides definitions and coding details for all variables used in the empirical analysis.

Source: PISA 2015, 2018, and 2022; CEPII; WVS.

Table A2: **Distribution of Destination and Origin Countries in the Analytical Sample**

Country	**Frequency**	**Percent (%)**	**Cumulative (%)**
Panel A. Destination Countries			
Australia	2,796	19.80	19.80
Austria	607	4.30	24.10
Belgium	220	1.56	25.66
Canada	2,906	20.58	46.25
Chile	164	1.16	47.41
Czech Republic	235	1.66	49.07
Denmark	307	2.17	51.25
Finland	902	6.39	57.64
Germany	137	0.97	58.61
Greece	232	1.64	60.25
Ireland	246	1.74	61.99
Luxembourg	1,660	11.76	73.75
New Zealand	1,635	11.58	85.33
Portugal	293	2.08	87.41
Switzerland	1,257	8.90	96.31
United Kingdom	521	3.69	100.00
Total (Destinations)	14,118	100.00	
Panel B. Origin Countries			
Afghanistan	127	0.90	0.90
Albania	269	1.91	2.80
United Arab Emirates	59	0.42	3.22
Australia	233	1.65	4.87
Austria	25	0.18	5.05
Bosnia and Herzegovina	32	0.23	5.28
Brazil	283	2.00	7.28
China	1,076	7.62	14.90
Colombia	43	0.30	15.21
Cape Verde	71	0.50	15.71
Germany	726	5.14	20.85
Spain	58	0.41	21.26
Estonia	269	1.91	23.17
Fiji	135	0.96	24.13
France	758	5.37	29.49
United Kingdom	1,568	11.11	40.60
Greece	29	0.21	40.81

Continued on next page

Table A2 Panel B. Origin Countries (continued)

Country	**Frequency**	**Percent (%)**	**Cumulative (%)**
Greenland	14	0.10	40.91
Croatia	109	0.77	41.68
Hungary	47	0.33	42.01
India	921	6.52	48.53
Ireland	379	2.68	51.22
Iran	88	0.62	51.84
Iraq	163	1.15	53.00
Iceland	22	0.16	53.15
Italy	458	3.24	56.40
Korea, Republic of	208	1.47	57.87
North Macedonia	15	0.11	57.98
Malaysia	49	0.35	58.32
Netherlands	17	0.12	58.44
Norway	11	0.08	58.52
New Zealand	503	3.56	62.08
Pakistan	245	1.74	63.82
Peru	21	0.15	63.97
Philippines	1,482	10.50	74.47
Poland	130	0.92	75.39
Portugal	1,253	8.88	84.26
Romania	94	0.67	84.93
Russia	279	1.98	86.90
Somalia	108	0.76	87.67
Slovakia	74	0.52	88.19
Sweden	73	0.52	88.71
Syria	229	1.62	90.33
Thailand	13	0.09	90.42
Tonga	133	0.94	91.37
Turkey	208	1.47	92.84
Ukraine	82	0.58	93.42
United States	331	2.34	95.76
Venezuela	100	0.71	96.47
Vietnam	175	1.24	97.71
South Africa	323	2.29	100.00
Total (Origins)	14,118	100.00	

Notes: Panel A reports the distribution of destination countries (current country of residence), while Panel B reports countries of origin. Percentages are calculated relative to the total analytical sample ($N = 14{,}118$). Country names follow ISO 3166–1 alpha–3 standards.

Source: PISA 2015, 2018 and 2022.

Table A3: **Categorical Distributions by Migration Status**

		Native students		Immigrant students	
Variable	**Category**	N	%	N	%
Panel A: Student characteristics					
Female	0 = Male	130,936	50.42	7,254	51.38
	1 = Female	128,761	49.58	6,864	48.62
Grade	7	490	0.19	92	0.65
	8	7,389	2.85	999	7.08
	9	69,213	26.65	3,700	26.21
	10	154,819	59.61	6,862	48.60
	11	25,795	9.93	2,370	16.79
	12	1,990	0.77	95	0.67
	13	3	0.00	–	–
Mother's education	0 = Lower primary	1,666	0.64	403	2.85
	1 = Primary	5,651	2.18	635	4.50
	2 = Lower secondary	35,698	13.75	1,899	13.45
	3 = Upper secondary	165,726	63.81	9,358	66.28
	4 = Tertiary/Vocational	50,958	19.62	1,823	12.91
Father's education	0 = Lower primary	2,615	1.01	390	2.76
	1 = Primary	8,091	3.12	732	5.18
	2 = Lower secondary	44,027	16.95	2,151	15.24
	3 = Upper secondary	145,292	55.95	8,960	63.47
	4 = Tertiary/Vocational	59,674	22.98	1,885	13.35
Books at home	1 = 0–10	21,120	8.13	2,068	14.65
	2 = 11–25	33,680	12.97	2,738	19.39
	3 = 26–100	65,681	25.29	3,498	24.78
	4 = 101–200	65,847	25.36	2,841	20.12
	5 = 201–500	46,148	17.77	1,801	12.76
	6 = 500+	27,223	10.48	1,172	8.30
Arrival age bin	Age 0–5	–	–	5,214	36.93
	Age 6–11	–	–	4,981	35.28
	Age 12–17	–	–	3,923	27.79
Panel B: School characteristics					
School location					
	1 = Village	530	6.33	116	4.43
	2 = Small town	1,842	21.99	436	16.64
	3 = Town	2,763	32.98	642	24.50
	4 = City	2,302	27.48	829	31.64
	5 = Large city	940	11.22	597	22.79
Public school					
	0 = Private	1,702	20.32	519	19.81
	1 = Public	6,675	79.68	2,101	80.19

Notes: This table reports distributions of key categorical variables for native and first-generation immigrant students, as well as school-level distributions for schools attended by immigrant students. Percentages are computed within migration group.

Source: PISA 2015, 2018, and 2022.

A.3 Additional Summary Statistics

Table A4: **Summary of Missing Values in Key Variables**

Variable	Missing	Total	% Missing
Panel A. Dependent Variables			
Integration	0	14,118	0.00
Bullying	0	13,448	4.74
Reading score	0	14,118	0.00
Panel B. Individual-Level Controls			
Age	0	14,118	0.00
Grade	0	14,118	0.00
Female	1	14,118	0.01
Mother's education	698	14,118	4.94
Father's education	991	14,118	7.02
Books at home	286	14,118	2.03
Panel C. School-Level Controls			
School location	0	2,620	0.00
Public school	0	2,620	0.00

Note: This table reports the extent of missing values for key variables used in the baseline analysis. Individual-level controls are aggregated to the school–year level. To preserve the full analytical sample and mitigate selection bias, indicator variables for missing values are included in all regressions.

Source: PISA 2015, 2018 and 2022.

Table A5: **Summary Statistics by Gender**

Variable	N (Male)	Mean (Male)	N (Female)	Mean (Female)	Mean diff.	t-stat
Integration	7,254	−0.029	6,864	−0.162	0.133	7.840***
Bullying	6,955	0.145	6,551	−0.075	0.219	11.559***
Reading	7,254	−0.233	6,864	0.007	−0.240	−13.396***

Notes: This table reports summary statistics by gender. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Source: PISA 2015, 2018 and 2022.

Table A6: **Summary Statistics for Students' Behavior and School Life**

Outcome	**N**	**Mean**	**SD**	**Min**	**Max**	**Wave**
Panel A: Educational and Occupational Expectations						
Expected years of education	12,697	14.948	2.613	9	20	2015, 2018, 2022
Expected occupational status	10,156	0.145	0.982	-2.647	1.342	2015, 2018, 2022
Expected occupational status relative to parents	10,156	-0.224	1.188	-4.410	2.837	2015, 2018, 2022
Panel B: Study Attitudes and Time						
Out-of-school study time (hours/week)	2,867	17.709	13.232	0	70	2015
Homework/study time at home	4,305	5.202	3.047	0	10	2015
Panel C: Teacher Relationships and School Climate						
Teacher unfairness	3,514	10.350	4.209	1	24	2015
Student–teacher relationship quality	3,208	0.020	1.043	-6.469	3.165	2022
Panel D: Socio-emotional Skills, Global Competence, and Psychological Traits						
Resilience	4,753	0.047	1.009	-3.168	2.369	2018
Attitudes toward immigrants	3,425	0.442	0.971	-2.264	1.499	2018
Cognitive flexibility	3,591	0.147	0.985	-3.278	2.145	2018
Respect for other cultures	3,496	0.230	0.886	-3.174	0.929	2018
Intercultural communication awareness	3,562	0.087	1.005	-2.795	2.051	2018
Global-mindedness	3,462	0.162	1.012	-2.878	2.634	2018
Empathy	4,066	0.020	1.040	-6.453	4.690	2022
Assertiveness	3,967	0.059	0.966	-8.211	7.230	2022
Emotional control	3,687	0.067	0.967	-5.170	5.533	2022
Growth mindset	3,958	0.203	1.042	-4.285	3.372	2022
Mathematics anxiety	3,607	0.098	1.152	-2.501	2.635	2022

Notes: This table reports summary statistics for students' behavior and school life. "Wave" denotes the PISA survey year(s) in which each outcome is available. Educational and occupational expectations are observed in 2015, 2018, and 2022. All continuous variables follow PISA's standardized scaling methodology.

Source: PISA 2015, 2018 and 2022

B IV: Construction and Correlation

B.1 Construction of the Composite Common Language Index

In empirical studies where language is not the central explanatory variable, but rather a background characteristic, it remains necessary to account for linguistic proximity across countries, as language similarity may confound cross-country comparisons. This issue is particularly relevant in country-level analyses with limited sample sizes, where the inclusion of multiple highly correlated language indicators can compromise estimation precision and identification.

To ensure a parsimonious specification, we rely on a composite measure of linguistic similarity that summarizes key language-related links without introducing unnecessary redundancy. This approach follows Melitz and Toubal (2014), who argue that certain language indicators convey overlapping information and therefore need not be included simultaneously. In particular, measures based on spoken language similarity are excluded, since their variation is largely subsumed by indicators capturing shared native language background.

The resulting index is designed to capture exogenous linguistic proximity and is bounded between zero and one. Its structure prioritizes native language similarity, which is treated as the most fundamental dimension of linguistic relatedness and therefore enters the index directly. Other language connections, including common official language status and deeper linguistic proximity, contribute only in cases where countries do not share a native language. This contribution is scaled to ensure that the index remains interpretable and bounded.

Formally, the index is defined as

$$\text{Common Language Index}_{ij} = \text{CNL}_{ij} + \frac{\text{COL}_{ij} + \text{LP2}_{ij}}{\max(\text{COL} + \text{LP2})} \cdot (1 - \text{CNL}_{ij}). \tag{8}$$

Under this construction, the index attains its maximum value when countries share a native language. In the absence of native language overlap, linguistic similarity is determined by the normalized contribution of official language links and genealogical language proximity.

By construction, this index provides a continuous and transparent measure of linguistic similarity that relies on predetermined language characteristics. Its compact form and exogeneity make it particularly suitable for instrumental variable applications in settings where language proficiency is potentially endogenous. Accordingly, we operationalize linguistic distance as one minus the similarity index (i.e., linguistic distance $= 1 -$ CLI), ensuring that higher values correspond to greater linguistic distance.

Table B1: **Language Measures Used in the Analysis**

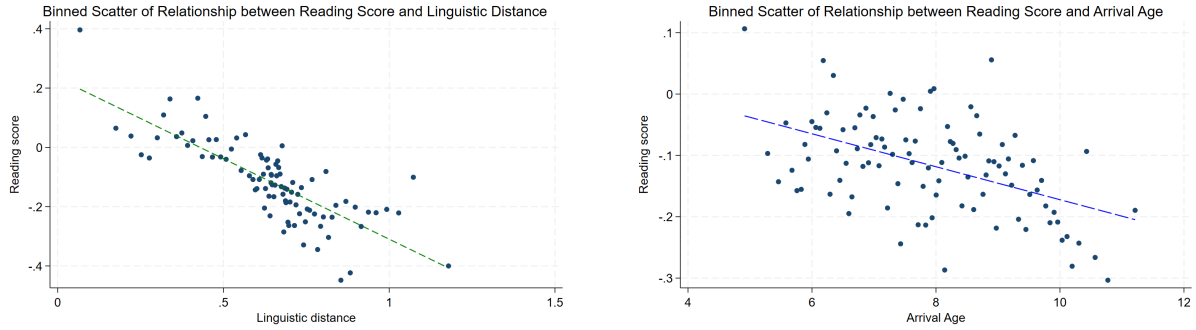
Variable	Definition	Scale	Measurement Description	Underlying Source
COL	Common Official Language	Dummy	Indicator equal to one when two countries designate at least one identical official language, independent of native usage or fluency.	Official language legislation and government records
CNL	Common Native Language	[0, 1]	Computed as the sum of cross-country products of population shares speaking the same language as their mother tongue.	Eurobarometer surveys and national population statistics
CSL	Common Spoken Language	[0, 1]	Constructed from population shares able to communicate in the same non-native language, with adjustments to avoid multiple counting across languages.	Eurobarometer data and multilingual survey sources
LP1	Linguistic Proximity I	> 0, mean $= 1$	Derived from genealogical similarity in the Ethnologue language tree, using discrete scores that reflect shared linguistic family structures.	Ethnologue classification and Fearon and Laitin (2003)
LP2	Linguistic Proximity II	> 0, mean $= 1$	Based on lexical similarity measures from the ASJP database using Swadesh word lists, transformed from distance into a proximity index.	ASJP project and Swadesh (1952)
CLI	Common Language Index	[0, 1]	Composite indicator aggregating information from COL, CNL, and linguistic proximity measures, excluding spoken language overlap.	Authors' calculations based on CEPII inputs

Notes:

- **Linguistic proximity measures:** LP1 captures categorical similarity based on linguistic family classifications, whereas LP2 reflects continuous lexical similarity. In most empirical applications, LP2 is preferred due to its richer variation.
- **Relationship between CNL and LP:** When two countries share a common native language, linguistic distance is not defined and the proximity measure is normalised to zero. When native languages differ, LP quantifies linguistic similarity based on structural or lexical characteristics.
- **Conceptual distinction between COL and CSL:** COL represents institutional alignment through official or educational language status, while CSL reflects population-level communicative exposure. The presence of a common official language does not necessarily imply widespread spoken proficiency.

Source: CEPII.

B.2 Relationship between Arrival Age, Linguistic Proximity and Language Proficiency



(a) **Linguistic Distance and Reading Score** (b) **Arrival Age and Reading Score**

Note: Panel (a) shows the relationship between linguistic distance and reading test scores, while Panel (b) shows the relationship between age at arrival and reading proficiency. The binned scatter plots residualize both variables with respect to year and school fixed effects. *Source:* PISA, CEPII.

Figure B1: **Relationship between Arrival Age, Linguistic Proximity and Language Proficiency**

Table B2: **Balance Tests on Predetermined Characteristics**

	Female	Mother education	Father education	Books
	(1)	(2)	(3)	(4)
Arrival age × Linguistic distance	0.001	-0.005	-0.002	0.004
	(0.003)	(0.004)	(0.004)	(0.008)
Controls	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	14,118	14,118	14,118	14,118

Notes: This table reports balance tests for the instrumental variable by regressing predetermined individual characteristics on the instrument (age at arrival × linguistic distance). All specifications include the same baseline controls and fixed effects as in the main analysis. Standard errors (in parentheses) are clustered at the school level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Source: PISA 2015, 2018, 2022; CEPII.

C Additional Main Results

C.1 Reduced Form

Table C1: **Reduced-Form Effects of the Instrument on Immigrant Students' Psychosocial Outcomes**

Model	Baseline RF (1)	Add Cultural Distance (2)	Origin FE (3)	Country–Origin Pair FE (4)	Country×Year FE (5)
Panel A: Dependent Variable – Integration					
Arrival age × Linguistic distance	-0.024***	-0.023***	-0.016***	-0.016***	-0.025***
	(0.006)	(0.006)	(0.005)	(0.005)	(0.006)
Controls	Yes	Yes	Yes	Yes	Yes
Cultural distance	No	Yes	No	No	No
Origin FE	No	No	Yes	No	No
Country-Origin FE	No	No	No	Yes	No
School FE	Yes	Yes	No	No	Yes
Year FE	Yes	Yes	Yes	Yes	No
Country×Year FE	No	No	No	No	Yes
N	14,118	14,118	14,118	14,118	14,118
Panel B: Dependent Variable – Bullying					
Arrival age × Linguistic distance	0.025***	0.024***	0.019***	0.018***	0.025***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.0075)
Controls	Yes	Yes	Yes	Yes	Yes
Cultural distance	No	Yes	No	No	No
Origin FE	No	No	Yes	No	No
Country-Origin FE	No	No	No	Yes	No
School FE	Yes	Yes	No	No	Yes
Year FE	Yes	Yes	Yes	Yes	No
Country×Year FE	No	No	No	No	Yes
N	13,448	13,448	13,448	13,448	13,448

Notes: This table reports reduced-form estimates of the effect of the instrument (age at arrival × linguistic distance) on psychosocial outcomes for first-generation immigrant students and corresponds to Table 2. Integration (Panel A) is a standardized index of school belonging, and bullying (Panel B) a standardized index of bullying exposure. All specifications control for age-at-arrival categories and continuous linguistic distance, include the same student- and school-level controls as the corresponding 2SLS specifications, and incorporate the fixed effects indicated. Standard errors (in parentheses) are clustered at the school level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Source: PISA 2015, 2018 and 2022; CEPII; WVS.

C.2 Main Regression: All Controls

Table C2: **2SLS Estimates for Reading Proficiency and Control Variables**

Variable	**Integration**	**Bullying**
	(1)	(2)
Second stage		
Reading proficiency	0.558***	-0.589***
	(0.153)	(0.181)
Arrival age 6–10	0.009	-0.041
	(0.027)	(0.026)
Arrival age 11–15	-0.042	-0.097*
	(0.056)	(0.060)
Linguistic distance	0.065	-0.123*
	(0.064)	(0.069)
Student characteristics		
Age	0.039	-0.023
	(0.038)	(0.039)
Grade	-0.134**	0.106
	(0.060)	(0.069)
Female	-0.246***	-0.076**
	(0.031)	(0.035)
Parent education		
Father education: primary	0.046	-0.116
	(0.091)	(0.121)
Father education: lower secondary	0.158*	-0.246**
	(0.088)	(0.125)
Father education: upper secondary	0.102	-0.153
	(0.096)	(0.127)
Father education: vocational	0.154*	-0.166
	(0.089)	(0.121)
Mother education: primary	0.065	-0.058
	(0.088)	(0.102)
Mother education: lower secondary	0.013	-0.089
	(0.078)	(0.099)
Mother education: upper secondary	-0.000	-0.119
	(0.084)	(0.105)
Mother education: vocational	0.025	-0.150
	(0.082)	(0.102)
Father education missing	0.084	-0.153**
	(0.063)	(0.074)
Mother education missing	-0.021	0.110
	(0.068)	(0.075)
Books at home		

Continued on next page

Variable	Integration	Bullying
	(1)	(2)
Books: 11–25	0.011	-0.003
	(0.051)	(0.054)
Books: 26–100	-0.069	0.141*
	(0.066)	(0.075)
Books: 101–200	-0.221**	0.231**
	(0.095)	(0.106)
Books: 201–500	-0.267**	0.363***
	(0.116)	(0.131)
Books: 500+	-0.339***	0.546***
	(0.100)	(0.120)
Books missing	-0.153**	0.400***
	(0.078)	(0.101)
School characteristics		
Small town	0.010	0.011
	(0.098)	(0.117)
Town	0.039	-0.154
	(0.095)	(0.111)
City	0.012	-0.105
	(0.100)	(0.117)
Large city	0.012	-0.076
	(0.110)	(0.125)
Public school	0.069	-0.078
	(0.065)	(0.068)
First stage		
Arrival age × linguistic distance	-0.043***	-0.042***
	(0.006)	(0.006)
KP F-statistic	69.67	60.23
School FE	Yes	Yes
Year FE	Yes	Yes
N	14118	13488

Note: This table reports full estimates of the effect of language proficiency on psychosocial outcomes for first-generation immigrant students. Integration is a standardized index of school belonging, and bullying a standardized index of bullying exposure. All specifications include student- and school-level covariates as well as school and year fixed effects. Standard errors (in parentheses) are clustered at the school level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Source PISA 2015, 2018 and 2022; CEPII.

C.3 Language Spoken at Home

Table C3: **Baseline Results: Speaking the Host-Country Language at Home and Outcomes of First-Generation Immigrant Students**

Dependent Variable	Integration		Bullying	
	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)
First Stage				
Arrival age × linguistic distance		-0.028***		-0.029***
		(0.002)		(0.002)
KP F-statistic		158.40		153.08
Second Stage				
Speak local language at home	0.075***	0.852***	-0.084***	-0.856***
	(0.024)	(0.218)	(0.027)	(0.248)
Controls	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
R^2	0.234	–	0.257	–
N	14,102	14,102	13,437	13,437

Notes: This table reports OLS and 2SLS estimates of the effect of speaking-language proficiency—measured by whether students speak the host-country language at home—on outcomes among first-generation immigrant students. Columns (1) and (3) report OLS estimates, while Columns (2) and (4) report 2SLS estimates, including first-stage results. The instrument is the interaction between age at arrival and linguistic distance. All specifications include student- and school-level characteristics and school and year fixed effects. Standard errors (in parentheses) are clustered at the school level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Source: PISA 2015, 2018 and 2022; CEPII.

D Additional Robustness Results

Table D1: **2SLS Results: Language Proficiency Effects by Country (Clusters ≥ 100)**

	First Stage			Second Stage			
Country	β_{IV}	SE	KP-F	$\beta_{\text{Reading score}}$	SE	N	School Clusters
Panel A. Integration							
Australia	-0.051***	0.012	20.27	0.483*	0.257	2,796	592
Austria	-0.009	0.024	0.47	-6.352	1313.632	607	163
Canada	-0.057***	0.015	19.85	0.346	0.246	2,906	540
Finland	-0.054**	0.026	4.90	0.865	0.647	902	172
New Zealand	-0.046***	0.016	11.40	0.391	0.326	1,635	188
Switzerland	-0.020	0.020	1.21	3.285	10.298	1,257	223
United Kingdom	-0.017	0.037	0.44	7.534	119.158	521	123
Panel B. Bullying							
Australia	-0.048***	0.012	17.22	-0.213	0.246	2,675	567
Austria	-0.012	0.024	0.67	-1.028	48.799	595	158
Canada	-0.058***	0.015	18.80	-0.405*	0.242	2,814	525
Finland	-0.053**	0.027	4.54	-1.225	0.954	872	165
New Zealand	-0.045***	0.017	9.33	-0.387	0.366	1,506	168
Switzerland	-0.020	0.023	0.96	-8.574	197.354	1,111	195
United Kingdom	-0.018	0.038	0.44	-0.092	5.270	512	121

Notes: This table reports country-specific 2SLS estimates of the effect of reading proficiency on school integration (Panel A) and bullying victimization (Panel B) among first-generation immigrant students. Reading proficiency is instrumented with age at arrival interacted with linguistic distance. All specifications include student- and school-level characteristics and absorb school and year fixed effects. Standard errors are clustered at the school level. KP-F reports the Kleibergen–Paap rk Wald F statistic (averaged across the ten plausible-value estimations). Only countries with at least 100 school clusters are reported to ensure reliable cluster-robust inference. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Source: PISA 2015, 2018, and 2022; CEPII.

Table D2: **Alternative Parameterizations of Arrival Age and Linguistic Distance**

	(1)	(2)	(3)	(4)	(5)
Panel A: Integration					
Second stage:					
Reading score	0.558***	0.451***	0.390**	0.335**	0.594***
	(0.153)	(0.125)	(0.167)	(0.134)	(0.080)
First stage:					
Arrival age × Linguistic distance	-0.043***	-0.038***	-0.049***	-0.039***	-0.039***
	(0.006)	(0.005)	(0.008)	(0.005)	(0.003)
KP F-statistic	69.67	91.17	49.86	68.55	262.10
Arrival age bins	Yes	Yes	No	No	No
Arrival age (continuous)	No	No	Yes	Yes	No
Linguistic distance (continuous)	Yes	No	Yes	No	No
Linguistic distance dummy (top quartile)	No	Yes	No	Yes	No
Controls	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
N	14,118	14,118	14,118	14,118	14,118
Panel B: Bullying					
Second stage:					
Reading score	-0.589***	-0.381***	-0.176	-0.112	-0.299***
	(0.181)	(0.139)	(0.194)	(0.150)	(0.086)
First stage:					
Arrival age × Linguistic distance	-0.042***	-0.038***	-0.048***	-0.038***	-0.039***
	(0.006)	(0.005)	(0.008)	(0.005)	(0.003)
KP F-statistic	60.23	82.63	42.80	61.96	240.16
Arrival age bins	Yes	Yes	No	No	No
Arrival age (continuous)	No	No	Yes	Yes	No
Linguistic distance (continuous)	Yes	No	Yes	No	No
Linguistic distance dummy (top quartile)	No	Yes	No	Yes	No
Controls	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
N	13,448	13,448	13,448	13,448	13,448

Notes: This table presents five specifications varying controls for arrival age and linguistic distance. Panel A reports 2SLS estimates for school integration, and Panel B for bullying. Column (1) reports the baseline specification; subsequent columns introduce alternative arrival-age and linguistic-distance controls. Arrival-age bins are defined as ages 1–5, 6–11, and 12–17 at arrival, and the linguistic-distance indicator equals one for students in the top quartile of the linguistic-distance distribution. All specifications include baseline student- and school-level characteristics and school and year fixed effects. Standard errors are clustered at the school level. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Source: PISA 2015, 2018, 2022; CEPII.