INVESTIGATING IMMIGRANT STUDENT
ACADEMIC ACHIEVEMENT ON PISA
BY LINKING ADDITIONAL DATA
ON ORIGIN CHARACTERISTICS

By

William Nicholas Bork Rodriguez

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Educational Psychology & Educational Technology—Doctor of Philosophy

2024

ABSTRACT

A secondary analysis of PISA 2018 data was conducted to investigate the reading achievement of students with an immigrant background. While prior research suggests the importance of demographic characteristics in secondary PISA analyses, a major critique is that PISA centers destination characteristics over origin ones, limiting the study of the association between origin characteristics and academic achievement. This study addressed the critique by linking PISA data with data of origin country characteristics to allow for analyses that could not be conducted with the PISA data set alone. The study linked PISA data with data of origin country characteristics, centered the origin-based characteristics in the analyses, and then evaluated the utility of linking this additional data for explaining education outcomes. Multilevel statistical modeling was used to model the association between origin country characteristics and the academic achievement of students with an immigrant background. Linked data were: (1) Language Distance—student-level measure of similarity between home and school language; (2) Human Development Index—country-level measure of human development; (3) Global Adaptation Index—country-level measure of climate vulnerability and readiness to improve resilience/climate adaptation; and (4) forced displacement ratio—country-level ratio between inward/outward forced displacement. The principal findings were the negative association between language distance and PISA reading scores (i.e., one unit increase in language distance associated with -0.31 unit change in reading score) and that language distance afforded a closer look at the association between language and PISA reading scores when applied to four particular countries and their language pairs (e.g., Switzerland, Finland, Qatar, Israel).

TABLE OF CONTENTS

INTRODUCTION

The Program for International Student Assessment (PISA) is an International Large-Scale Assessments (ILSA) that assesses the knowledge and skills of 15-year old students and reports results in domain areas such as mathematics, reading, and science (OECD, 2019a). ILSAs are empirical studies to assess educational achievement (Rocher & Hastedt, 2020). The results inform stakeholders, such as educators, researchers, policymakers, and the general public about the status of learner achievement and their educational system, which often guide decisions regarding the evaluation and reform of education at scale (Rocher & Hastedt, 2020).

Students with an immigrant background are one important group that participate in PISA. Prior research suggests that immigrant student achievement on PISA is typically studied using secondary analyses and that there are numerous benefits in doing so (Donnellan et al., 2011; Torney-Purta & Amadeo, 2013a). Prior research also suggests that demographic characteristics have been a typical research focus for PISA secondary analyses (Aloisi & Tymms, 2017). In recent years, PISA results have shown mixed achievement amongst immigrant students (Schleicher, 2006; OECD, 2016). In addition, PISA results have shown increasing immigration in most countries (OECD, 2019a).

While prior research suggests the importance of demographic characteristics, a major critique of this research is that PISA centers destination characteristics over origin ones in studies. This is because PISA questionnaire items collect data from students who have since settled into their post-immigration destination country (i.e., destination country). Therefore, most of the data is about students' lived experiences within the

destination country, with a much smaller number of items on characteristics related to participants' origin country (e.g., parents' country of birth). This limits researchers from studying how characteristics of the origin country may introduce variation into the academic achievement of students with an immigrant background.

One way to address the critique of research is by linking PISA data with data of origin country characteristics to enhance the analytic opportunities by allowing for analyses that could not be conducted with the PISA data set alone. A study was proposed to this end. The study was built on three pillars. The first pillar was linking PISA data with data of origin country characteristics to enhance the analytic opportunities, allowing for analyses that could not be conducted with the PISA data set alone. The second pillar was to center the origin-based characteristics in the analyses. The third pillar was to evaluate the utility of linking this additional data for explaining education outcomes. The study linked PISA data with four additional data sets bringing origin country characteristics into the analysis. The study used multilevel statistical modeling to model the association between origin country characteristics and the academic achievement of students with an immigrant background.

PHENOMENON & RATIONALE OF THE STUDY

The Phenomenon of Study: Reading Achievement of Immigrant Students on the Programme for International Student Assessment (PISA)

This study investigated the phenomenon of reading achievement of students with an immigrant background on the Programme for International Student Assessment (PISA), a large-scale international assessment (ILSA).

Rationale for the Importance of Studying This Phenomenon

The rationale for studying this phenomenon came from a review of research which suggested two important reasons for studying this phenomenon. First, there is variation in immigrant student reading outcomes by their individual and country characteristics (OECD, 2016). Studying these characteristics can help explain and predict immigrant student achievement. Second, there is an increasing number of students with an immigrant background in most countries (OECD, 2019b). This increase suggests the growing importance regarding the attention and resources directed towards educating learners with an immigrant background.

REVIEW OF RESEARCH

The review of prior research highlighted topics relevant to this study. To start, an overview of the PISA assessment is given. Then the first relevant review topic highlighted that immigrant student achievement on PISA has typically been studied using secondary analyses. Second was that demographic characteristics has been a typical research focus for PISA secondary analyses. Third was that PISA results have shown mixed achievement amongst immigrant students. Fourth was that PISA results have shown increasing immigration in most countries.

Then a critique of the prior research was highlighted: while prior research suggests the importance of demographic characteristics, a critique of this research is that PISA centers destination characteristics over origin ones in studies. This limits researchers from studying how characteristics of the origin country may introduce variation into the academic achievement of students with an immigrant background.

Finally, an assertion was made that one way to address the critique of research was by linking PISA data with data of origin country characteristics to enhance the analytic opportunities by allowing for analyses that could not be conducted with the PISA data set alone.

An Overview of the PISA Assessment

Large-Scale International Assessments (ILSAs) have a history spanning over twenty years, starting with the introduction of the Trends in International Mathematics and Science Study (TIMSS) in 1995 to the most recent Global Teaching Insights (GTI) study in 2020 (Rocher & Hastedt, 2020). Appendix A lists prominent ILSAs. ILSAs typically assess key knowledge and skills and report results in domain areas such as

mathematics, reading, science, and more. This study focused on the ILSA reading achievement of students with an immigrant background because language-based skills are typically, though not always, more challenging for immigrant students to develop compared to more technical domains such as mathematics and science (Schnepf, 2006; Andon et al., 2014).

One of the most well known ILSA is the Programme for International Student Assessment (PISA). PISA was first introduced in 2000 by the Organisation for Economic Co-operation and Development (OECD). The rationale for PISA was to meet a need for internationally comparable evidence on student performance, to offer insights for education policy and practice, and to monitor trends in acquisition of knowledge and skills (OECD, 2019a). The research questions of PISA, at a high level, ask to what extent have 15-year-old students acquired key knowledge and skills essential for full participation in social and economic life (OECD, 2019a).

PISA instrumentation is a computer-based assessment of four 30-minute clusters of assessment material presented to students (OECD, 2019c). The clusters were drawn from mathematics, science, and reading domains (OECD, 2019c). Test items included both multiple-choice and constructed-response questions (OECD, 2019c). Additionally, contextual data was collected from 35-minute long surveys administered separately to students, their teachers, and their school principals (OECD, 2019c).

PISA methods are complex, with affordances and constraints different from smaller scoped assessments. Specific sample designs, weights, and plausible values are used, typical to other ILSAs (Bailey et al., 2023). For example, PISA is a balanced-incomplete-block (BIB) design that assesses reading, mathematics, science,

and global competence (OECD, 2019c). BIB designs group test items into clusters which are presented to participants with equal frequency and position across tests (OECD, 2019c). Approximately 600,000 students participated in 79 countries/economies for PISA 2018 (OECD, 2019c). Participants were selected based on a two-stage stratified sampling procedure where first schools were sampled and then students were sampled within the selected schools (OECD, 2019c). Participant performance was reported on a continuous scale based on item-response theory models, as well as, arranged into six numbered proficiency levels (i.e., 6 is highest) to help users interpret PISA scores (OECD, 2019c).

Immigrant Student Achievement Typically Studied with Secondary Analyses

The first review topic was that immigrant student achievement on PISA has typically been studied using secondary analyses. This is because PISA is designed to support secondary analyses with publicly available data and ample supporting documentation. Prior research suggests there are numerous benefits in conducting secondary analysis of immigrant student achievement on PISA (Donnellan et al., 2011; Torney-Purta & Amadeo, 2013a). There are multiple reasons why secondary analyses of ILSAs are important. Torney-Purta & Amadeo (2013a) wrote that "...careful secondary analysis can address important research questions about human development, learning and education at both the macro and micro levels" (p. 249). In addition, quantitative secondary analysis of ILSAs can generate research questions to be answered by smaller mixed-methods studies (Torney-Purta & Amadeo, 2013a). Torney-Purta & Amadeo (2013a) stated that secondary analyses can be especially impactful in the educational and developmental psychology field where data are typically limited in

sample size and scope. Secondary analyses give researchers the potential to make inferences at multiple levels of analysis, ranging from the macro level (e.g., understanding country-level education outcomes), down to the micro level (e.g., classroom practices), and including the levels between (Torney-Purta & Amadeo, 2013b). Secondary analyses also promote an open source approach to science by making the artifacts and processes of scientific research (e.g., data, analyses, etc) broadly accessible to actors outside the study (Donnellan et al., 2011). This saves time and costs on data collection, aids reproducibility of findings, and promotes careful documentation of processes (Donnellan et al., 2011).

Demographic Characteristics are a Typical Research Focus for PISA Secondary Analyses

The second review topic was that demographic characteristics have been a typical research focus as PISA surveys collect lots of demographic information. Survey questions typically ask students about their home environment, their school environment, and even questions about how their parents engage with the world around them. One example of a study making use of demographic characteristics data comes from Cattaneo and Wolter (2015) with their investigation into how country policy efforts impacted immigrant student outcomes. This study controlled for demographic characteristics first, to then attribute the remaining variation to recent changes in migration policy in Switzerland. The rationale for the Cattaneo and Wolter (2015) study was that in Switzerland, between 2000 and 2009, improvements in socio-economic background were traditionally attributed to integration policies. However, migration policy may have also had an impact on socio-economic background as well (Cattaneo &

Wolter, 2015). The research question asked about the extent to which achievement improvements in 1st-generation immigrant students were attributable to changes in observable background variables due to migration policy changes? The methods used linear regression to model reading scores for PISA 2000 and PISA 2009. The authors' key explanatory variables came from PISA and included parents' socio-economic background characteristics such as an SES index, parents' education, books in home, parental occupation, home language, and parent nationality (Cattaneo & Wolter, 2015). In addition to PISA survey data, the study incorporated additional data from the parental occupation index from International Socio-Economic Index of Occupational Status (Cattaneo & Wolter, 2015). One finding of the study was that 1st generation immigrants scored 40-points higher on PISA between 2000 and 2009 (Cattaneo & Wolter, 2015). The authors reported that 75% of that score increase was attributed to changes in parent background characteristics and school composition while the remaining 25% was attributed to changes in migration policy (e.g., immigration policy and law).

Other researchers have found demographic characteristics useful for studying the academic outcomes of immigrant students. A study by Aloisi and Tymms (2017) found that demographic factors are more impactful than policy and reform ones. The rationale for the Aloisi and Tymms (2017) study was to see if adjusting for factors outside the control of education policymakers could better distill out the impact of policy. The researchers asked three research questions: (1) What is the relationship between changes in the socio-economic and demographic characteristics of the PISA cohorts, and changes in country outcomes? (2) What is the relationship between changes in the curricular provision of PISA-participating countries and their outcomes? (3) Overall,

what is the relative importance of non-policy-malleable factors (student SES and demographics), when compared with policy-malleable factors (e.g., curricular changes) with respect to PISA scores? The researchers used multilevel growth models. In addition to PISA survey data between PISA 2000 and PISA 2015, additional external data was incorporated for country reforms information (Aloisi & Tymms, 2017). The authors found that demographic characteristics of students (e.g., SES level) impacted PISA achievement more than reform policies (Aloisi & Tymms, 2017). In fact, the study reported an annual effect size of only 0.02 standard deviations for reform policies (Aloisi & Tymms, 2017). The study did not find strong evidence for the effectiveness of curricular reforms on PISA outcomes. The authors concluded that education reforms typically change the education goal or vision but the less malleable demographic factors explain more variation in PISA scores (Aloisi & Tymms, 2017).

PISA Results Suggest Mixed Achievement Amongst Immigrant Students

Another review topic was the mixed achievement amongst immigrant students' PISA scores. One early report of immigrant learner outcomes came from a comparative review of PISA 2003 data in which Schleicher (2006) reported that not all immigrant PISA participants scored alike. The Schleicher (2006) report was commissioned to examine performance differences between students with and without an immigrant background and to identify what impacted the results, so as to provide implications for educational policy. One group of research questions in the report asked about an immigrant student performance gap with sub-questions for economic, social, and cultural background characteristics while the other group of questions asked about between-country differences and potential policy intervention points (Schleicher, 2006).

The analysis was conducted using multiple regression for continuous outcome variables and logistic regression for dichotomous outcome variables and no unique instruments were used as the data were already collected for PISA 2003 (Schleicher, 2006). One finding was that immigrant students performed lower than native learners, with variation by country (Schleicher, 2006). A second finding was there was no significant association between immigrant population size and performance outcomes (Schleicher, 2006). A third finding was that background characteristics of immigrant students, after controlling for them, only partially explained the variation in performance outcomes in some countries (Schleicher, 2006). A fourth finding was that differences in student language and the language of instruction also partially explained the variation in performance outcomes, with variation by country (Schleicher, 2006). A final finding was that countries with well-established language support programs featuring clear goals and standards tended to have immigrant learners who perform more similarly to native learners (Schleicher, 2006).

In more recent times, the official PISA 2015 results have also reported variation in immigrant student academic outcomes and demographic characteristics (OECD, 2016). One finding was the 1.4% increase in overall educated parents and student SES between PISA 2009 and PISA 2015 (OECD, 2016). There were 57.3% of first-generation immigrant students with a parent who had an education level equal to the average non-immigrant parent (OECD, 2016). Countries with a 10+% increase were: Belgium, Croatia, Denmark, Ireland, and Luxembourg. Relatedly, some immigrant students had a similar or higher economic, social, and cultural status (ESCS) than their non-immigrant peers within: Estonia, Ireland, Latvia, Malta, Montenegro, Singapore, and

United Arab Emirates (OECD, 2016). A second finding was increased differences between home and school languages (OECD, 2016). There was a 10+% increased difference in: Belgium, Germany, Greece, Ireland, Qatar, and Slovenia (OECD, 2016). Among immigrant students, those who did not speak the PISA test language at home scored 54 points lower than non-immigrant students while those who did speak the PISA test language at home scored relatively better at only 31 points lower (OECD, 2016). This trend was true for all subjects but less pronounced in mathematics with only a 15 point deficit (OECD, 2016). The countries with the highest language penalty were: Hong Kong, Luxembourg, Austria, Belgium, Jordan, Macao, Russia, and Switzerland (OECD, 2016). A third finding was that immigrant learners scored 31 points lower in science compared to non-immigrant learners, after controlling for SES (OECD, 2016). The largest deficits (i.e., 40 to 55 points) were in: Austria, Belgium, Denmark, Germany, Slovenia, Sweden, and Switzerland (OECD, 2016). In other countries, the scores were similar between immigrant and non-immigrant students who both scored above the OECD average in: Australia, Canada, Estonia, Hong Kong, Ireland, and New Zealand (OECD, 2016). Furthermore, in some countries, immigrant learners outperformed non-immigrants with the largest differences (i.e., 22 to 80 points) found in: Macao, Qatar, and the United Arab Emirates (OECD, 2016). A fourth finding was that SES disadvantage only partially accounted for immigrant learner outcomes (OECD, 2016). In 22 of the 33 countries with at least a 6.25% immigrant student population, there were significant differences in science performance between immigrant and non-immigrant learners, after controlling for socio-economic status (OECD, 2016). Only in 5 countries did the immigrant background effect disappear after controlling for socio-economic

11

status in: Costa Rica, Hong Kong, Israel, Singapore, and the United States. Similar

results were reported for mathematics and reading (OECD, 2016). A fifth finding was

that 1st-generation immigrant learners performed better in culturally similar destinations

compared to their peers in culturally different ones, after controlling for similar SES

(OECD, 2016). For example, 1st-generation mainland Chinese students who moved to

culturally similar locations (e.g., Hong Kong or Macao) scored better than those who

moved to culturally different ones (e.g., Australia or New Zealand) (OECD, 2016).

However, the cultural familiarity effect was not found for 2nd-generation immigrants

(OECD, 2016). For example, 2nd-generation mainland Chinese students scored better

in culturally different destinations (e.g., Australia or New Zealand) than those

2nd-generation mainland Chinese students in culturally similar ones (e.g., Hong Kong or

Macao) (OECD, 2016).

PISA Results Suggest Increasing Immigration in Most Countries

Another review topic was that PISA has detected an increase in the number of

students with an immigrant background over time (OECD, 2019b). PISA 2018 reported

one finding that the OECD country average of students with an immigrant background

had increased from 10% to 13% between PISA 2009 and PISA 2018 (OECD, 2019b).

The largest increases were seen in: Canada, Ireland, Luxembourg, Malta, Norway,

Qatar, Singapore, Sweden, Switzerland, and the United Kingdom (OECD, 2019b). A

second finding was an increase in socio-economically disadvantaged students as

measured by the bottom quartile of PISA's economic, social, and cultural status (ESCS)

index (OECD, 2019b). At least 45% of immigrant students were disadvantaged in:

Austria, Denmark, Finland, France, Germany, Greece, Iceland, the Netherlands,

Norway, Slovenia, and Sweden (OECD, 2019b). Conversely, some other countries have gained immigrant students from higher socio-economic status; higher than their own non-immigrant population (OECD, 2019b). This suggested the migration of highly skilled workers in places such as: Brunei Darussalam, Panama, Qatar, Saudi Arabia, Singapore, and the United Arab Emirates (OECD, 2019b). A third finding was that home and school language differences were common with 48% of immigrant students not speaking the PISA assessment language while at home (OECD, 2019b). This threshold was over 70% in: Austria, Brunei Darussalam, Finland, Iceland, Lebanon, Luxembourg, and Slovenia (OECD, 2019b). Conversely, this value was <10% in: Costa Rica, Croatia, Jordan, and Kazakhstan (OECD, 2019b). Speaking the assessment language at home was associated with increased performance in: Brunei, Darussalam, Germany, Luxembourg, Macao, Malta, and Switzerland (OECD, 2019b). A fourth finding was that there were distinct effects of immigration status that were not explained by SES (OECD, 2019b). Generally, non-immigrant students outperform first- and second-generation immigrant students, when controlling for students' and schools' SES (OECD, 2019b). Furthermore, immigrant students with the same origin country and SES could still perform differently within different destination countries (OECD, 2019b). For example, immigrant students had a 30+ point reading deficit in: Austria, Denmark, Estonia, Finland, Iceland, Lebanon, Norway, and Sweden (OECD, 2019b). Conversely, some immigrant students performed better than non-immigrant peers in: Australia, Brunei Darussalam, Hong Kong, Jordan, Macao, Qatar, Saudi Arabia, the United Arab Emirates, and the United States (OECD, 2019b).

The Major Critique of Prior Research: Secondary PISA Research on Immigrant

Students Centers Destination Characteristics over Origin Ones

While prior research suggests the importance of demographic characteristics, a critique of this research is that PISA centers destination characteristics over origin ones in studies. This is because PISA questionnaire items collect data from students who have since settled into their post-immigration destination country (i.e., destination country). Therefore, most of the data is about students' lived experiences within the destination country, with a much smaller number of items on characteristics related to participants' origin country (e.g., parents' country of birth). This is a reasonable design choice given the goals of PISA. However, this means that PISA can provide little data about the life of students and their parents prior to immigration, which in turn limits researchers from studying how characteristics of the origin country may introduce variation into the academic achievement of students with an immigrant background. Table 1 shows a breakdown of demographic question counts by origin country or destination country for the PISA 2018 survey. A third column is for demographic questions that overlap both origin and destination questions. See Appendix B for a more detailed breakdown of the count by specific questionnaire (i.e., student, teacher, or school questionnaires).

Table 1

Count of Demographic Questions on PISA 2018

Origin Demographics	Destination Demographics	Both
1 question	15 questions	7 questions

Addressing a Critique by Linking PISA with Data of Origin Country Characteristics

While destination country data is useful, incorporating more origin country characteristics may provide a more complete picture of the immigrant student experience. Some researchers have linked ILSA data with additional data sets to enhance the analytic opportunities. This allows for analysis that could not be conducted with a single ILSA data set alone.

Affordances of Data Linking: Linking ILSAs with Additional Data Enhances Analytic Opportunities

Data linking studies have been an effective approach for studying immigrant outcomes in a number of social sciences. One such example is linking data sets to study health outcomes of people with an immigrant background. One example study by Giuntella et al. (2018) linked immigration circumstances with health outcomes. The rationale for the study was to investigate the relationship between reasons for immigration and health outcomes in the United Kingdom. The research questions asked about health outcomes for four different immigration reasons: employment, family, study, and asylum. The study used linear regression models, accounting for age, education, gender, ethnicity, residence, and year. Data was collected from the UK quarterly Labor Force Survey. A main finding was that immigrants who immigrated for employment, study, and family reported better health outcomes than native-born individuals (Giuntella et al., 2018). In addition, asylum seekers reported worse health outcomes (Giuntella et al., 2018).

Another example is linking data sets to study economic outcomes of people with an immigrant background. A study by Maskileyson et al. (2021) linked immigration circumstances with economic outcomes. The rationale for the study was to investigate the economic advantages of immigrants to Switzerland by their immigration circumstance (i.e., economical, political, family reunion, or educational pursuit). The research question asked if economic immigrants to Switzerland had higher income attainment compared to immigrants arriving for other reasons. The study used regression analysis of an existing data set. Data was collected for the 2007 Swiss health Survey. One major finding was that immigration motive was a significant factor in income attainment and that the results varied by gender (Maskileyson et al., 2021). Overall, economic reasons for immigration resulted in the highest income levels followed by educational reasons (Maskileyson et al., 2021). In addition, women tended to attain lower incomes than men (Maskileyson et al., 2021).

A Minor Critique of Prior Data Linking Studies: Lagging Adoption of Studies Linking Origin Data with Education Outcomes

Despite the affordances that data linking can provide for enhancing an analysis, a minor critique of prior data linking studies was regarding the lagging adoption of studies linking origin data with education outcomes. Instead, this type of research is more common in other social science fields such as in studying health or economic outcomes of people with an immigrant background. The lagging adoption in education means there is less prior research to rely upon when planning related studies on education outcomes. This leaves a research gap that, if filled, helps to evaluate the

utility of linking additional origin country data for explaining the academic achievement

for immigrant students on PISA (see Figure 1).

Figure 1

Research with Linked Data
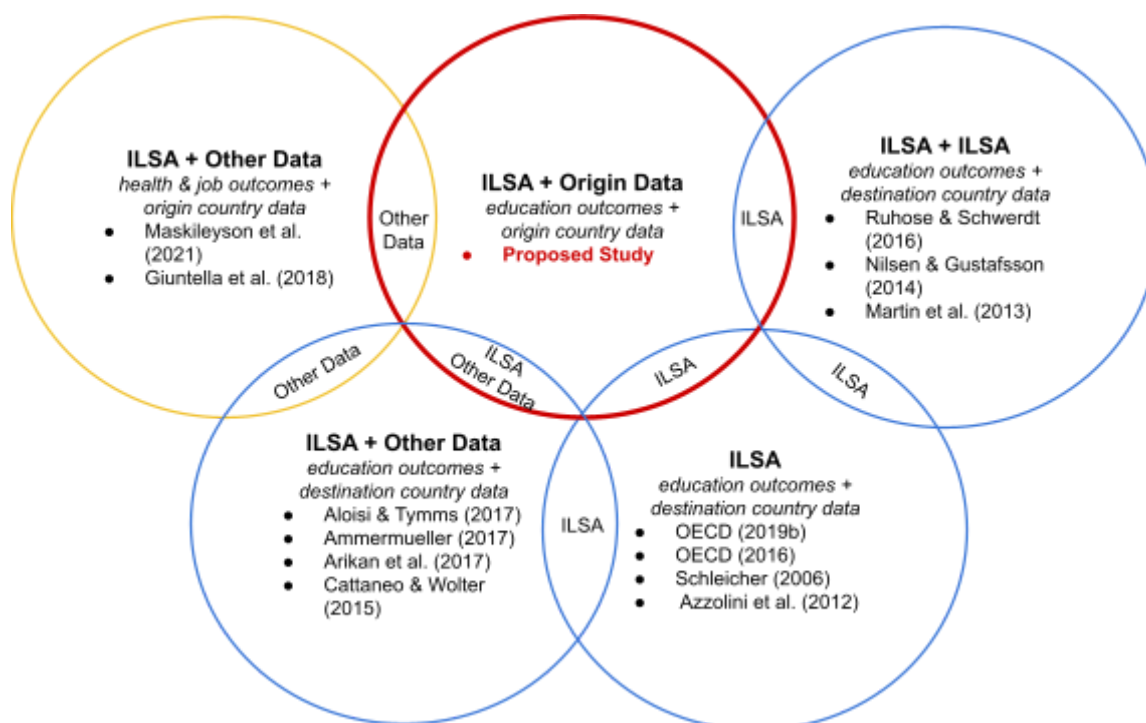
Fortunately, despite the research gap, there is a blueprint to follow for linking ILSA data,

such as PISA, with additional data sources to enhance the analysis. Specifically, other

researchers have provided a set of rationals and a framework for doing these types of

studies.

There Are Three Rationales for Linking Outside Data Sets

While ILSAs are designed to be stand-alone studies, some researchers have

linked additional data sources, from outside the ILSA, to enhance the analytic

opportunities of the study. An analysis of PISA data in conjunction with non-PISA data

may better explain the association between origin country characteristics and the

achievement of immigrant learners. There is already precedent for doing these types of data linking studies. Strietholt and Scherer (2016) provided three rationales. One rationale is to research changes over time. For example, Nilsen and Gustafsson (2014) linked data from two consecutive administrations of the same ILSA. The rationale for the study was to illustrate how different cycles of the same ILSAs can be combined at the test item level. The research question asked if changes in Norwegian schools' emphasis on academic skills caused an increase in science performance. The methods combined TIMSS 2007 and TIMSS 2011 (i.e., student and teacher/classroom data) on anchor items. The analysis was conducted using multilevel structural equation models. The study instruments were two prior implementations of TIMSS. One finding was that schools' emphasis on academic success was associated with science performance (Nilsen & Gustafsson, 2014).

A second rationale is to extend outcome measures. For example, Martin et al. (2013) linked two different ILSAs to be able to say more about outcomes than a single ILSA could say. The rationale for the study was that no ILSA simultaneously assesses reading, math, and science at once. The research questions asked questions about the effects of home environment on students' achievement. The study's methods linked TIMSS 2011 data (math and science) with PIRLS 2011 data (reading). Additionally, context data for home, school, and classroom contexts were combined for both ILSAs. The data were linked using unique identification numbers of students who took both ILSAs in 2011. The analysis used multilevel regression. The instruments were one TIMSS and one PIRLS implementation. One finding was that the results were stable for

the three combined academic domains of reading, math, and science, with country-level variation in how school characteristics were related to achievement (Martin et al., 2013).

A third rationale is to supplement outcome measures with context information. For example, Ruhose & Schwerdt (2016) linked ILSA data with information about academic tracking. The rationale for the study was that the researchers wanted to study the impact of early tracking of students (i.e., ability grouping) using UNESCO data on school structure by country. The researchers asked research questions about the effect of tracking on migrant and non-migrant students in various countries. The methods linked achievement data from PIRLS, PISA, and TIMSS with supplemental data on the country-specific age that tracking begins, from the UNESCO International Bureau of Education. The instruments of the study were the 3 ILSAs. The analyses used a difference-in-difference strategy. The authors' major finding was that early tracking did not impact achievement gaps between migrant and non-migrant learners (Ruhose & Schwerdt, 2016).

There Is A Framework to Guide the Linking of Data Sets

In addition to the rationales for data linking provided by Strietholt and Scherer (2016), the authors also provided suggestions for evaluating data attachment points, based on a framework by Bray and Thomas (1995) for comparative and multilevel analyses in educational studies (see Figure 2). The framework consists of three dimensions. The first dimension is geographic/locational (Bray & Thomas, 1995). There are seven levels in this dimension: world regions/continents, countries, states/provinces, districts, schools, classrooms, and individuals. The second dimension is nonlocational demographic groupings (Bray & Thomas, 1995). Some groups in this dimension

include: ethnicity, religion, age, gender, etc. The third dimension is regarding aspects of

education and society (Bray & Thomas, 1995). Some of the aspects in this dimension

include: curriculum, teaching methods, finance, management structures, political

change, and labor markets.

Figure 2

Attachment Points for Linking Data Sets
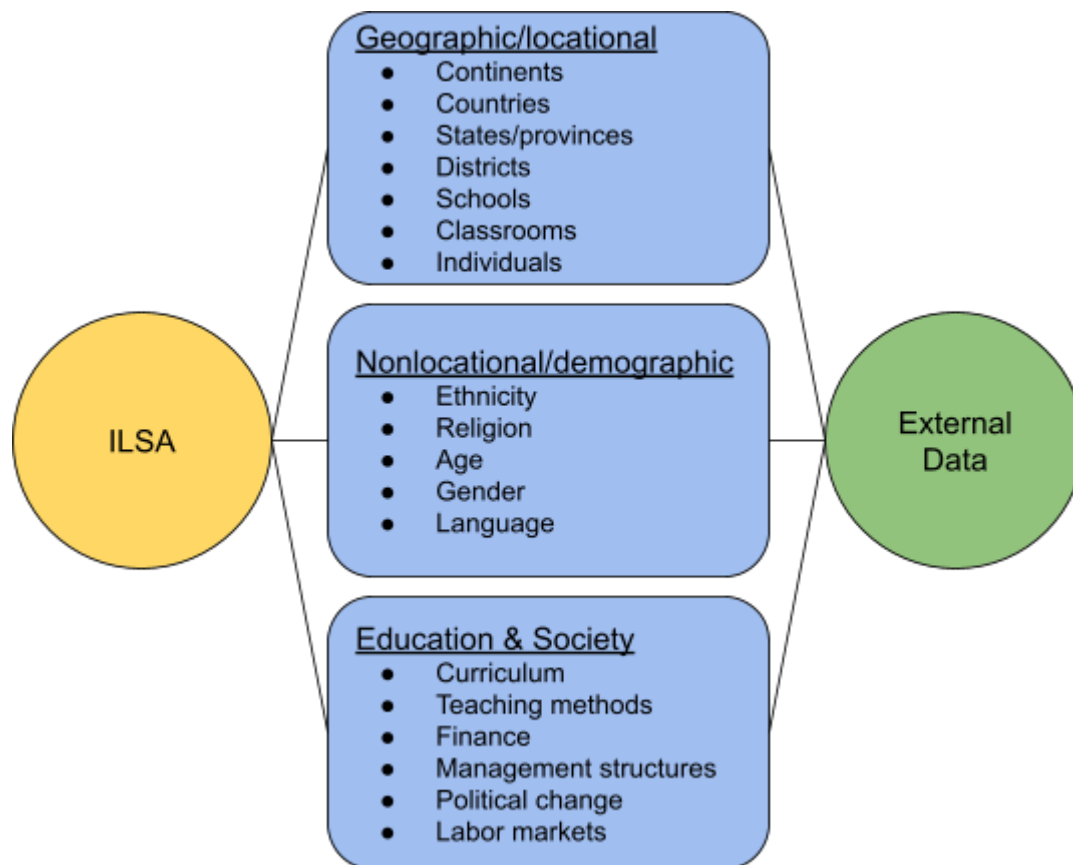
One illustrative example of the Bray and Thomas (1995) framework is seen in

Arikan et al. (2017) who linked data at three different attachment points: (1)

geographic/locational; (2) nonlocational demographic grouping; and (3) education and

society. Arikan et al. (2017) linked data external with ILSA data as a way to measure the

context of immigrant learners. The rationale for the Arikan et al. (2017) study was the

observed variation in the ILSA performance of Turkish immigrant students (i.e.,

nonlocational demographic grouping) by their destination countries (i.e.,

geographic/locational). The authors attempted to understand Turkish immigrant

performance as a function of both individual- and country-level characteristics across

multiple countries. The research question asked about the achievement differences

across countries and how they related to national integration policies (i.e., education

and society). The study's method was multilevel regression analysis of factors predicting

reading and mathematics performance. Predictor variables included both

individual-level (e.g., immigration status, social status, etc.) and country-level variables.

The country-level variables (e.g., longevity, health, knowledge, standard of living, etc.)

were incorporated from the Human Development Index (HDI). Additional country-level

data incorporated from the Migrant Integration Policy Index (MIPEX) provided measures

of anti-discrimination and general integration of migrant learners. The study instruments

were PISA 2009 data with linked country-level external data. One finding was that at the

individual-level, immigrant learners performed lower than their mainstream peers in

reading and math performance, even when controlling for an index of economic, social,

and cultural status (Arikan et al., 2017). Overall, students higher in economic, social,

and cultural status performed higher as well (Arikan et al., 2017). Another finding was

that at the country-level, destination countries with higher human development index

scores were associated with higher reading performance, though not math (Arikan et

al., 2017).

Turning specifically towards PISA, PISA data provides some attachment points

for researchers to link additional data. For instance, the questionnaire includes

questions about students' and parents' birth country. This provides a point of attachment for linking at the country-level. Another example is the student's age of immigration. This provides a time frame in which relevant origin country data should be linked.

Summary of the Research Review & Critique Points

The review of prior research highlighted topics relevant to this study. To start, an overview of the PISA assessment was given. Then the first relevant review topic highlighted that immigrant student achievement on PISA has typically been studied using secondary analyses and that there are numerous benefits in doing so (Donnellan et al., 2011; Torney-Purta & Amadeo, 2013a). Second was that demographic characteristics have been a typical research focus for PISA secondary analyses. Demographic characteristics of students (e.g., SES level) impacted PISA achievement more than reform policies (Aloisi & Tymms, 2017). While education reforms typically change the education goal or vision, it is the less malleable demographic factors that explain more variation in PISA scores (Aloisi & Tymms, 2017). Third was that PISA results have shown mixed achievement amongst immigrant students. PISA results suggest that not all immigrant PISA participants scored alike (Schleicher, 2006). PISA reported variation in immigrant student academic outcomes and demographic characteristics (OECD, 2016). Fourth was the increasing immigration in most countries. PISA has detected an increase in the number of students with an immigrant background over time (OECD, 2019a).

Then the major critique of the prior research was highlighted: while prior research suggests the importance of demographic characteristics, a critique of this research is that PISA centers destination characteristics over origin ones in studies. This is because

PISA questionnaire items collect data from students who have since settled into their post-immigration destination country (i.e., destination country). Therefore, most of the data is about students' lived experiences within the destination country, with a much smaller number of items on characteristics related to participants' origin country (e.g., parents' country of birth). This limits researchers from studying how characteristics of the origin country may introduce variation into the academic achievement of students with an immigrant background.

Finally, the study makes an assertion that one way to address the critique of research was by linking PISA data with data of origin country characteristics to enhance the analytic opportunities by allowing for analyses that could not be conducted with the PISA data set alone. A minor critique of prior data linking studies was stated, that despite the affordances that data linking can provide for enhancing an analysis, there is less prior research to rely upon when planning related studies on education outcomes. This leaves a research gap that, if filled, helps to evaluate the utility of linking additional origin country data for explaining the academic achievement for immigrant students on PISA. Fortunately, despite the research gap, there is a blueprint to follow for linking ILSA data, such as PISA, with additional data sources to enhance the analysis. Specifically, other researchers have provided a set of rationals and a framework for doing these types of studies. Strietholt and Scherer (2016) provided three rationales such as changes over time (Nilsen & Gustafsson, 2014), extending outcome measures (Martin et al., 2013), or supplement outcome measures with context information (Ruhose & Schwerdt, 2016). The framework suggests three points of attachment which

were geographical/locational, nonlocational demographic groupings, and aspects of

education and society (Bray & Thomas, 1995; Strietholt & Scherer, 2016).

THE PROPOSED STUDY

The Proposed Study: A Multilevel Statistical Analysis of the Association Between Origin Characteristics and the Academic Achievement of Students with an Immigrant Background

The review of research highlighted a major critique and made an assertion on how to address this critique. These aspects of the review of literature served as the foundation for the proposed research study. The study was built on three pillars (see Figure 3). The first pillar was linking PISA data with data of origin country characteristics to enhance the analytic opportunities, allowing for analyses that could not be conducted with the PISA data set alone. The second pillar was to center the origin-based characteristics in the analyses. The third pillar was to evaluate the utility of linking this additional data for explaining education outcomes. The study linked PISA data with four additional data sets bringing origin country characteristics into the analysis. The study used multilevel statistical modeling to model the association between origin country characteristics and the academic achievement of students with an immigrant background.

Figure 3

Pillars of the Proposed Study
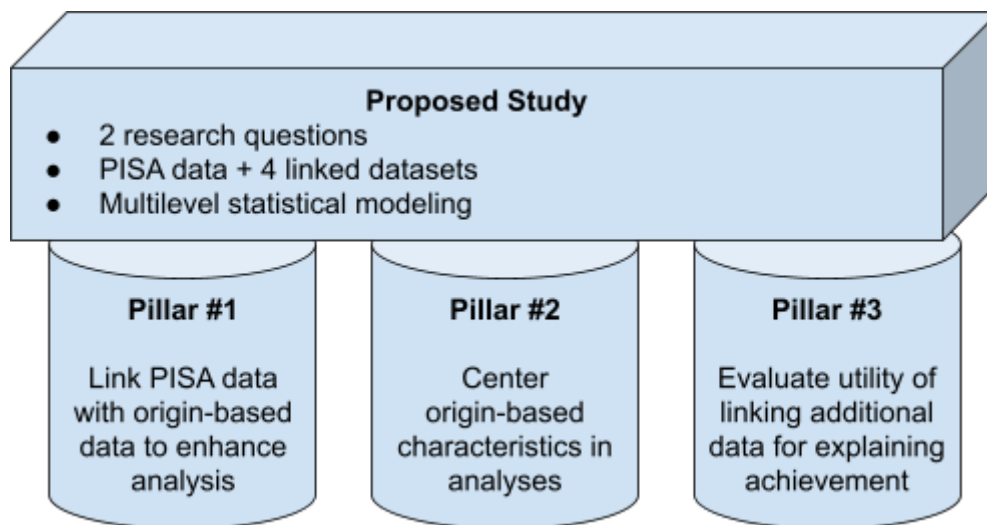
Research Questions Were Aligned to Pillars of the Study

This study posed two research questions which were aligned to the pillars of the proposed study:

1. Which specific origin country characteristics from the linked data sets have statistical significance for interpreting immigrant students' PISA reading achievement?

2. How much additional variation in immigrant students' PISA reading achievement is explained by the linked data sets?

Research question #1 was aligned to the first two pillars. These pillars were about bringing in additional data related to origin-based characteristics and making them the focus of the statistical analysis. The answers to research question #1 provided answers as to whether the selected data sets enhance the analysis.

Research question #2 was aligned with the third pillar. This pillar was about evaluating the assertion made in this study. The answers to research question #2

provided answers as to whether the additional analyses afforded by the linked data sets are worth the effort of linking them in the first place.

The Primary Data Are Sourced from PISA 2018; The Analytic Sample Was Selected by Immigration Status

PISA 2018 was the primary data source. The data set, as it comes from PISA, includes 546,932 students from 77 countries or territories. The participants of this study were selected into the analytic sample by their immigration status. Immigrant students numbered 14,246 participants (2.6% of all students). The PISA assessment categorizes participants by their immigration status. Participants are categorized as immigrants when their mother and father were born in a country different from the country where the PISA assessment was taken (OECD, 2016). Participants are categorized as non-immigrant when either their mother or father were born in the same country where the PISA assessment was taken OECD (2016). While PISA does differentiate between first-generation and second-generation immigrants, this study focused on just 1st generation immigrant students. There are two reasons for this. One reason was that the PISA 2018 sample did not contain substantial numbers of second generation immigrant students. The second reason was that working with second generation students would add complexity to the study. This would be due to uncertainty around the non-immigrant side of the family's familiarity with the destination country. This would confound the analysis in a way that is not present when working with just first generation immigrant students. In the end, the analytic sample was further reduced due to missing language data resulting in 9493 students with an immigrant background, 3514 schools, 42 destination countries, and 74 origin countries.

The Linked Data Were Sourced from Four Publicly Available Data Sets Using Four

Criteria

The selection of data sources was based on four criteria. First, the data set must

provide characteristics related to immigrant students' origin country. The reason for this

is that the study focuses on characteristics that were imparted by the origin country.

Second, the data set must have data for the years in which a student has been alive,

plus some years prior. The reason for this is to ensure there is data available for the

relevant years leading up to the point of immigration and beyond. Third, the data set

must be openly published. One reason for this was to make it easier to evaluate the

available data sets freely without having to request access from a data broker. An

additional reason for this is to allow future researchers to reproduce and extend on this

study without needing to obtain data permissions. Fourth, the data sets must provide

information that is not already measured by the ILSA. The reason for this was because

of the study's focus on bringing in new data not already available in PISA.

This search criteria yielded 4 data sets which are listed in Table 2. Further details

on these data sets are explained in the paragraphs below. Each data set contributed a

variable of interest which was used in the statistical modeling. The Procedures section

of this research report provides details on the work done with each individual data set.

Table 2

Data Sources for Immigration Circumstance

Name	Source	Description	Contribution
Language Distance	Automated Similarity Judgment Program (ASJP) via the Max Planck Institute for the Science of Human History	Matrix measure of the degree of linguistic similarity between languages in origin and destination countries.	Student-level language characteristics imparted by origin/destination countries
Human Development Index	United Nations Development Program (UNDP)	Composite measure of country-level human development based on three dimensions: health, education, and standard of living.	Origin country-level development characteristics
Global Adaptation Index	University of Notre Dame (ND-GAIN)	Index measure of country-level climate vulnerability and readiness to improve resilience/climate adaptation	Origin country-level climate characteristics
Refugee Population Statistics Database	United Nations High Commissioner for Refugees (UNHCR)	Count measures of the population of forcibly displaced persons in origin and destination countries.	Origin/Destination country-level forced displacement

Language Distance (LD)

The academic achievement of immigrant learners may be a function of both characteristics of their origin country, their destination country, and how the two interact. An illustrative example of this dynamic is language use and one way to use language analytically is with linguistic distance. Language distance is a measure of how different one language is from another language (Chiswick & Miller, 2005; Gamallo et al., 2017). A few researchers have created and validated their own quantitative measures of language distance (Chiswick & Miller, 2005; Gamallo et al., 2017; Bakker et al., 2009) or developed tools based on that research (Wichmann et al., 2022). Language distance can help explain some of the variation of immigrant learners' educational experience. An early hypothesis was that the less distance between origin and destination country languages, the faster and overall higher attainment level is reached (Corder, 1981). Seminal research by Heath and Heath (1983) provided evidence that learners who grow up outside the dominant culture don't learn the dominant culture's ways of doing discourse, which has negative effects on participation within education spaces. Furthermore, language distance can also influence immigration destination choices. For example, a study by Chiswick and Miller (1994) found immigrants to Canada were more likely to live in Quebec, an area where the romance language French is used, if they were arriving from a country that also used a romance language. Additionally, those romance language immigrants were also more likely to become French speakers after arrival (Chiswick & Miller, 1994).

Prior Research with Language Distance. Some researchers have used language distance to help explain the immigrant experience in a few ways. Beenstock et

al. (2001) studied the Hebrew language proficiency of immigrants to Israel. The rationale for the study was to separately assess the effects of origin country and origin language on destination language proficiency of immigrants. The research question asked whether immigrant language proficiency is an effect of country characteristics or origin language characteristics. The study's methods used an ordered logistic regression analysis. Data on Hebrew language fluency and literacy was linked from two data sources: a census data set and an immigrant absorption survey (Beenstock et al., 2001). One finding was that greater linguistic distance was associated with greater difficulty of learning the destination language (Beenstock et al., 2001). Arabic speakers had the smallest linguistic distance to Hebrew and thus Arabic speaking immigrants were the most proficient with Hebrew (Beenstock et al., 2001). Additionally, immigrants who originated from multilingual countries exhibited higher Hebrew proficiency as well (Beenstock et al., 2001).

In more recent times, other researchers have used measures of linguistic distance to study student academic achievement. Figueiredo et al. (2016) stated that heritage language speakers often struggled in European classrooms. The research questions asked about the effect of age and home language on the development of verbal reasoning and vocabulary tasks in second language learners. The methods started with a sample of 106 Portuguese language learners. Regression analysis was conducted. Data were collected using tests of reading, writing, and comprehension skills. Results showed home language was a significant predictor of variation in learners' outcomes with linguistic distance explaining the relationships (Figueiredo et al., 2016). Regarding linguistic distance from the target language of Portuguese,

31

learners coming from Indo-Aryan languages struggled the most while those coming from Romance languages struggled less (Figueiredo et al., 2016).

A study by Jain (2017) used language distance to investigate how the 1956 reorganization of South Indian states along strict linguistic lines affected academic achievement. The study's rationale was that matching between students' native language and school language may have influenced long-term academic achievement outcomes. The study asked two research questions. First, if a language mismatch benefits or hinders long-term education achievement. Second, whether alignment between home and school language leads to catch-up achievement. The study used linear regression analysis. Data was sourced from Indian census records ten years apart starting from 1951 through to 1991. An additional data set was sourced from a different study to provide data on population characteristics. One finding was that mismatched home and school language was associated with achievement (Jain, 2017). A second finding was that language-based reorganization of state boundaries may have remediated achievement gaps, with the greatest impact in areas with a longer history of language mismatch (Jain, 2017).

Why is Language Distance Important? Language distance is important because a smaller language distance can have positive results for learners as they can more quickly and deeply access the educational benefits of the destination country. A study by Ammermueller (2007) provided evidence that a match between home and school language is associated with higher achievement. The rationale of the Ammermueller (2007) study was to follow-up on the German PISA 2000 results that reported large differences between immigrant and non-immigrant students. The

research question asked: Why did immigrant and non-immigrant students achieve so differently on PISA 2000? The study's methods used linear regression and a Juhn-Murphy-Pierce decomposition method. In addition to PISA 2000 survey data, the study incorporated additional data from a Germany-specific PISA extension study. One finding was that parents' choice to use the destination language as the home language explained higher achievement (Ammermueller, 2007). Additionally, achievement could also be partially explained by enrolling earlier in German schools, thus gaining additional familiarity with the destination country's schooling culture (Ammermueller, 2007).

When a language match between home and school languages cannot be achieved, there is evidence that a partial language match also provides benefits. One study that investigated this was Azzolini et al. (2012). The rationale for the Azzolini et al. (2012) study was to investigate if immigrant student achievement was similar in both the traditional and recently popular immigration destinations. The researchers' questions asked: What is the variation in math and reading achievement by immigrant status in Italy and Spain? How much of the variation is accounted for by family background? The study's methods were hierarchical linear modeling at the student- and school-level fitted with predictor variables for family socioeconomic background, home language, and school characteristics. The Azzolini et al. (2012) analysis did not implement unique instruments as it relied on PISA 2009 survey data. One finding was that students with a home language different from the test language score lower (Azzolini et al., 2012). However, the gap was smaller in Spain when the immigrant student spoke Latin American Spanish, compared to Italy where a language match was less common

(Azzolini et al., 2012). These trends were similar in both traditional and newer immigration destinations (Azzolini et al., 2012). Another finding was the variation in achievement was partially attributed to parent occupation and economic integration (Azzolini et al., 2012).

Language distance is important for the present study because it informs on how the origin country language may influence pre-immigration decisions and post-immigration communication.

Language Distance is Linked to PISA at the Student-Level. A measure of language distance comes from the Automated Similarity Judgment Program (ASJP), created by researchers from the Max Planck Institute for Evolutionary Anthropology. ASJP is used to calculate a lexical distance between languages using the Levenshtein Distance algorithm (1966) and a database of Swadesh Word Lists (Swadish, 1955). The latest versions of software and tools (Wichmann et al., 2022) were developed based on research from Bakker et al. (2009). ASJP produces matrices of language distance between pairs of languages (Wichmann et al., 2022). Language distances are provided in decimal form with lower values indicating less distance between languages and larger values indicating greater language distance (Wichmann et al., 2022). For instance, the language distance between German and English is 67.00 while German and Spanish is 92.98; indicating that German is more similar to English than to Spanish (Wichmann et al., 2022). These values can be used to indicate the similarity or dissimilarity between an immigrant students' origin country language and their destination country language.

These data are linked at the student-level, producing a language distance value for each participant in the PISA data. It is important to note that the language distance

measure differs from the other measures used in this study in one key aspect. Language distance is a student-level measure while the other measures are country-level measures. Language distance represents language characteristics at the student-level, which were imparted by the origin and destination countries. In other words, it reflects characteristics of the student's origin country carried over into the destination country. For example, the language a student uses in the home is associated with their country of origin. Therefore, a student's language distance values are influenced by the origin country.

Human Development Index (HDI)

The Human Development Index (HDI) is a composite measure of country-level human development. It is based on three dimensions: health, education, and standard of living. Health is assessed by life expectancy (UNDP, 2023). Education is measured by years of schooling (UNDP, 2023). Standard of living is measured by the logarithm of gross national income per capita (UNDP, 2023). The aim of HDI is to provide a measure of how people and their capabilities, not just economics, can be used to assess the development of a country (UNDP, 2023). HDI was developed with a new concept of what human development should mean; the process of enlarging the set of choices available to people (Klugman et al., 2011). HDI was developed by a Pakistani economist named Mahbub ul Haq and collaborating scholars including Amartya Sen who wanted a different way of measuring development with the then typical per capita income measure (Klugman et al., 2011). HDI is a relatively simple and transparent measure, which has facilitated its wide adoption since its inception in the year 1993 (Klugman et al., 2011). Since then, HDI has often been used by governments for informing policy or

resource allocation decisions (Dervis & Klugman, 2011). However, HDI is not without

criticisms. One critique is that HDI should broaden to include constructs such as gender

equity (Dervis & Klugman, 2011; Sharma, 2010). A different critique is that some

countries have large distribution gaps in their indicator levels (e.g., life expectancy)

between subgroups (Ghislandi et al., 2019). Another critique is that HDI indicators don't

allow much discrimination between the countries at the highest and lowest ends of the

distributions (Kovacevic, 2010).

Prior Research with HDI? Within research, HDI is typically used as a country-level

variable in studies where development level is suspected to be an influential factor in

outcome measures. For instance, a study by Vargas-Montoya et al. (2023) investigated

the relationship between the use of information and communications technologies (ICT)

and student learning outcomes. The Vargas-Montoya et al. (2023) definition of ICT

included the use of computers to search or practice skills. The rationale for the study

was the inconclusive evidence on whether the use of ICT was associated with positive

learning outcomes. The researchers asked questions about whether the country context

was relevant to the association. The study's methods utilized hierarchical linear

modeling with HDI as one measure of country context to measure whether, and how,

development levels impacted the relationship. The instruments included PISA 2018 data

and HDI data. One result was that this relationship did indeed differ depending on

whether students were in a country deemed developed or developing (Vargas-Montoya

et al., 2023). The authors concluded that country context is an important consideration

as results can vary significantly enough for countries by their categorical development

level (Vargas-Montoya et al., 2023).

Why is HDI Important? An important contribution of HDI was that it decoupled human development from a strictly economic measure, such as income and Gross Domestic Product measures (Dervis & Klugman, 2011; Klugman et al, 2011). HDI is an important index because of its wide acceptance, having gained an audience with political leaders, scholars, and activists (Dervis & Klugman, 2011). This means that studies which use the HDI can readily integrate into the broader conversation around country-level human development (Dervis & Klugman, 2011; Klugman et al, 2011). HDI is important for the present study because it provides a measure of country-level context related to indicators that may hold influence on immigration decisions (i.e., health, education, and standard of living).

HDI is Linked to PISA Data at the Country-Level. These data are linked at the country-level, producing an HDI value for each participant in the PISA data.

Global Adaptation Index (GAIN)

The Notre Dame-Global Adaptation Initiative annually provides the free and open source Notre Dame-Global Adaptation Index (GAIN). The GAIN index includes 45 indicators across 181 countries to report a combined country-level measure of climate vulnerability and readiness to improve resilience/climate adaptation (Chen et al., 2015). The aim of GAIN is to help governments, businesses, and communities use the data to prioritize investments for climate adaptation and lowering the risk of climate vulnerability (Chen et al., 2015).

Prior Research with GAIN. GAIN has been used in a variety of quantitative studies to investigate how climate change impacts a variety of outcomes. Some examples of scholarly work that used the GAIN have included work on climate

adaptation actions in agriculture (Nowak & Rosentock, 2020) and climate-induced vulnerability in the water sector (Nyiwul, 2023). A study by Chen et al. (2018) investigated the equity and efficiency of allocation of resources for climate change adaptation. A study by Werrell et al., (2015) investigated how climate conditions contributed to societal instability in Syria and Egypt which lead up to 2011 Arab Spring; a series of anti-government protests, uprisings and rebellions.

Why is GAIN Important? The GAIN index is important because it provides not only an assessment of country-level vulnerability to climate change, but it also provides information on the readiness for making adaptations to mitigate risk (Chen et al., 2015). This makes the GAIN index both present-focused and future-focused (Chen et al., 2015). Present day vulnerabilities to climate changes are reflected in the vulnerability indicators (e.g., food, water, health, ecosystem service, human habitat, and infrastructure) (Chen et al., 2015). Future ability to endure climate change is reflected in the readiness indicators (e.g., economic readiness, governance readiness, and social readiness) (Chen et al., 2015). This is important since the timetable of climate change has not progressed linearly (Chen et al., 2015). GAIN is important to the present study because it reflects factors that may guide decisions to immigrate. Specifically, when residents of a country look around and see indicators of the country's readiness or lack of readiness for future climate change, this may influence decisions to immigrate before the full effects of climate change have been realized.

GAIN is Linked to PISA Data at the Country-Level. These data are linked at the country-level, producing an GAIN value for each participant in the PISA data.

Forced Displacement (FD): Refugee Population Statistics Database

Data on forced displacement is typically used to study human migration flows, specifically of vulnerable groups such as displaced persons, asylum seekers, refugees, stateless people, etc. (UNHCR, 2023). While there are numerous data sets of forced displacement, they don't all cover the same groups of people, with some focusing on more regional or more national levels. The data for this present study comes from the United Nations High Commissioner for Refugees (UNHCR) Refugee Population Statistics Database (UNHCR, 2023). This database provides data on forcibly displaced populations which include refugees, asylum-seekers, and internally displaced people, and stateless persons (UNHCR, 2023). Forced displacement has been described as an emerging development challenge with extreme poverty increasingly concentrated among vulnerable groups such as those who are fleeing conflict and violence (World Bank, 2017).

Prior Research with Forced Displacement. Some researchers have used data on forced displacement in research on refugee experiences and refugee impacts on asylum countries. Education outcomes are one subcategory of that refugee experience. A report by Dryden-Peterson (2015) was written with the rationale to address a literature gap for post-arrival refugee education experiences in the United States of America. The research asked questions regarding which elements of refugee students' education were relevant to their post-resettlement education. The Dryden-Peterson (2015) methods included an analysis of three data sources on refugee education experience: access and quality of education for refugees, field-based case studies, and key informant interviews. The instruments were existing UNHCR data for refugee education

access and quality, in addition to the author's own fieldwork. Select findings were that: refugee learners typically face disrupted schooling, resulting in skill and knowledge gaps; pre-resettlement schooling is typically sporadic; and refugee learners are exposed to multiple languages of instruction (Dryden-Peterson, 2015).

Why is Forced Displacement Important? This data is important because the data allow researchers to make country-level inferences about the status of immigration climate and policy regarding immigrants in general (World Bank, 2017). Additionally, it provides information on both country of origin and destination, allowing researchers to follow country-level movement. Furthermore, displaced people can face xenophobia from the communities in which they are hosted (World Bank, 2017). The forced displacement data is important to the present study because it provides a high level sense of the degree to which underlying factors may have driven immigration, both forced and unforced.

Forced Displacement is Linked to PISA Data at the Country-Level. The UNHCR reports data annually on country-level forced displacement. The forced displacement data used in this study was obtained from the World Bank Open Data website, measured by the refugee population by country or territory of origin and the refugee population by country or territory of asylum. These data are linked on the country-level, specifically connecting via PISA participants' destination and origin country. In this study, these data are operationalized as a forced displacement ratio called FD Ratio. FD Ratio is a country-level ratio between inward forced displacement to outward forced displacement. For instance, if 2000 forcibly displaced persons entered a country while 1000 people were forced out by that same country, then the ratio is

2000:1000, simplified to 2:1, and 2.0 in decimal form. Decimal form is used for statistical

modeling in this study. The decimal values ranged from 0.0 to around 12,000 at the high

end. Therefore, a value of 12,000 means that there are 12,000 forcibly displaced

persons arriving for every 1 person going out. Higher ratios tend to be more

wealthy/desirable countries (i.e., many arrive, few leave). Examples include Denmark,

Sweden, and Norway. Lower ratio tends to be less wealthy/desirable countries (i.e., few

arrive, many leave). Examples include The Philippines, Afghanistan, and Albania).

Acknowledging the Causality Limitations of ILSAs

A limitation of ILSAs is that they are cross-sectional studies, which means they

collect data on a particular sample of students at one point in time. In contrast are

longitudinal studies which collect pre- and post- data from the same sample of students.

A critique of ILSAs is that they are not longitudinal and therefore cannot control for prior

levels of a given measure. Therefore, ILSAs are not able to establish a causal link

between explanatory variables and an outcome variable as they lack the pre- and

post-test design.

However, these constraints can be mitigated in two ways. One way is to use

findings from related, yet different studies to estimate an expected effect that a prior

achievement measure would have on regression outcomes. An example of this can be

seen in a study by Carnoy et al. (2016) who tried to improve ILSAs results by

incorporating longitudinal data to provide more precise, less biased estimated

estimates. Carnoy et al. (2016) critiqued the use of ILSAs for making education policy

recommendations as cross-sectional studies can identify broad trends but are not able

to provide causal inferences. The authors contended that this leads to simplified and

misleading conclusions, compared to true longitudinal studies. Therefore, their study was intended to reduce the bias in typical estimates obtained from cross-sectional ILSAs (Carnoy et al., 2016). The researchers asked if researchers could obtain more accurate estimates of the effects of classroom variables on students' PISA performance. The study used linear regression models to estimate PISA 2012 math achievement using a stepwise sequence of explanatory variables such as individual student characteristics; class and school characteristics; and teacher characteristics. The study instruments were two ILSA data sets; TIMSS 2011 and PISA 2012, both from Russia. The same sample of students took both these assessments one year apart, allowing the researchers to control for prior academic achievement in a way that either ILSA could not provide alone (Carnoy et al., 2016). The major methodological finding for this study was that controlling for previous math achievement resulted in more modest estimates for the relationships between explanatory variables and PISA outcome scores (Carnoy et al., 2016). Before the researchers included prior achievement (i.e., TIMSS 2011 scores) into their models, the effect sizes of pre-service teacher training on PISA math scores ranged from -0.16 to -0.21 standard deviations (Carnoy et al., 2016). After controlling for prior achievement, the estimates reduced to a range of -0.14 to -0.16 standard deviations (Carnoy et al., 2016). Therefore, controlling for prior achievement may lessen the effect size of PISA outcomes between 0.02 and 0.05 standard deviations (Carnoy et al., 2016).

Another example comes from Jerrim et al. (2022) who studied the effectiveness of inquiry-based science teaching by linking ILSA data with other measures of science attainment. The rationale for the study was the ongoing debate on the effectiveness of

42

the widely used inquiry-based science teaching in England. There were two research questions: (1) Do young people who receive a higher frequency of inquiry-based science teaching have higher levels of science achievement? and (2) Is there a positive association between specific components of inquiry-based teaching and young people's achievement in science? Jerrim et al. (2022) ran regression models to estimate science achievement on General Certificate of Secondary Education (GCSE) scores, controlling for demographic, prior achievement, and school-level measures. The main outcome variable was GCSE science achievement with a secondary analysis using PISA science attainment as an alternative outcome. The instruments of the Jerrim et al. (2022) study included two main data sources: (1) England's PISA 2015 and (2) England's 2016 National Pupil Database (NPD). Data was linked across these data sets. This provided measures of prior achievement at age 11 (Key Stage 2), age 15 (PISA), and age 16 (GCSE). The main finding was that inquiry-based teaching had a weak and inconsistent relationship with GCSE science attainment (Jerrim et al., 2022). From the methodological perspective, when the researchers included prior achievement (Key Stage 2 scores) into their models this reduced the effect sizes of inquiry-teaching on PISA science scores, by 0.02, 0.03, and 0.12 standard deviations for the respective quartiles (Jerrim et al., 2022). Adding in all the remaining controls further reduced effects sizes by an additional 0.04, 0.03, and 0.00 standard deviations (Jerrim et al., 2022). Therefore, controlling for prior achievement may lessen the effect size of PISA outcomes between 0.02 to 0.12 standard deviations (Jerrim et al., 2022).

An alternative way to address causality in ILSAs is to forgo accounting for prior achievement altogether, as controlling for prior achievement may obscure the

cumulative negative effects of systemic inequity on academic achievement.

Researchers working from a critical-quantitative approach (i.e., Quantcrit) might argue against the assumption of the obligate inclusion of some traditional controls, such as prior achievement (Sablan, 2019). Controlling for prior achievement may obscure the cumulative negative effects of systemic inequity on academic achievement (Frank et al., 2023a). Take for example an intervention study measuring the effect of some teaching practice. In this example, a pre- and post-test of academic achievement is conducted. The resulting analysis then shows that the teaching effect increased achievement by 0.5 standard deviations. From a typical quantitative perspective, this suggests the effectiveness of the teaching practice. However, since the study only measured achievement between the pre-test and the post-test, the researcher may overlook that one group of students (e.g. students of a particular race) still scored significantly lower than other groups, despite similar growth in the time between the pre- and post-test (Frank et al., 2023a; Sablan, 2019). In other words, while all students improved at a similar rate due to the intervention, the preexisting achievement gaps between groups still persists (Frank et al., 2023a; Sablan, 2019). Therefore, omitting pre-tests of prior achievement allows the effects of negative systemic factors to be more obvious (Frank et al., 2023a; Sablan, 2019).

The above paragraphs provided two different approaches for addressing the limitation of ILSAs as cross-sectional studies. One approach is to rely on studies that have approximated the pre- and post-test designs with other data sources (Jerrim et al., 2022). The other approach is a purposeful omission of pre-tests (Frank et al., 2023a; Sablan, 2019). Both approaches are relevant to this study. On the one hand, this study

can run models that incorporate a proxy measure of prior achievement based on related studies. On the other hand, the rationale for omitting controls of prior achievement is relevant to the study as it investigates how immigrant students' experience of society may impact their academic achievement rather than identifying specific causal elements. Taken together, the variation between the different approaches may provide helpful context to the analyses.

METHODS

One Dependent Variable Was Sourced from PISA 2018: Reading Scores

The outcome variable in the analysis was the student-level reading scores on PISA 2018. The PISA assessment reports scores as plausible values (PV). Plausible values represent a distribution of possible scores for a given participant, rather than a single individual test score (OECD, 2009). Plausible values represent a range of the scores an individual student would be reasonably expected to score, based on their actual point estimates and the associated probability of these values (OECD, 2009). Plausible values are designed to reduce measurement bias occurring from using a relatively small number of items that a given PISA participant actually sees (OECD, 2009). Typically, modeling software runs regression for each PV; averages the runs (e.g., 10 PVs = 10 models). However, not all software handles plausible values in every case. In particular one software used in this study does not handle plausible values when modeling cross-nested multilevel models. This software was the Scientific Software International (SSI) HLM software (Raudenbush & Congdon, 2021). Not using plausible values introduces bias into estimates and is a limitation of the study. On the one hand, not using plausible values biases estimates towards more extreme scores (Von Davier et al., 2009). On the other hand, averaging PVs biases towards underestimated group-level variances (Von Davier et al., 2009). This study's compromise is to use just a single plausible value when utilizing software with PV limitations. Overall, this approach is interpreted as providing less confidence in smaller score differences (Von Davier et al., 2009).

Four Explanatory Variables Were Sourced From The Linked Data Sets

The outcome variable in the study was a student-level reading PISA score. Explanatory variables include a number of demographic variables that come from PISA as well as the variables from the non-PISA data sets which were linked to the PISA data. Additionally, the variables include operational variables needed to conduct the analysis such as weighting variables and anonymized student identity numbers. The complete list of variables is found in Appendix C.

Materials: The Complexity of the Data Called for Multiple Tools

There are multiple software tools used for this study. Together, these software tools were used to create an analytic sample (i.e., R, Tidyverse, EdSurvey, Countrycode, and fastDummies) and run statistical models (i.e., SSI HLM, WeMix). Table 3 provides more detail about the software tools used in this study. A few other software options were considered in this study on the basis of their functionality for running HLMs. A comparison table is shown in Appendix D.

Table 3

Software used for Data Prep and Statistical Models

Software	Purpose
R (v. 4.2.1)	A programming language and environment for statistical computing and graphics (R Core Team, 2023).
Tidyverse (v. 2.0.0)	A collection of R packages designed for data science, with a focus on data import, tidying, manipulation, visualization, and programming (Wickham et al., 2019).
EdSurvey (v. 3.1.0)	A collection of functions for working with complex sample designs, weights, and plausible values common to NCES, OECD, and IEA education survey and assessment data (Bailey et al., 2023)
Countrycode (v. 1.3.0)	An R package to convert between country names and coding schemes (Arel-Bundock et al., 2018).
fastDummies (v. 1.7.1)	The goal of fastDummies is to quickly create dummy variables (columns) and dummy rows (Kaplan & Schlegel, 2023)
WeMix (v. 4.0.0)	A collection of functions for conducting HLMs with multilevel data that includes weights at multiple levels (Bailey, et al., 2021).
SSI HLM (v. 8.2)	A specially designed collection of modeling tools for modeling with a wide variety with hierarchical data (Raudenbush & Congdon, 2021).

Procedures: Data Preparation Was Needed Prior to Linking Data Sets

Obtained PISA 2018 Data

The first step was to obtain the ILSA data. The PISA 2018 data was obtained

from the OECD PISA 2018 data set website. The website provided a series of

downloadable data files. The data download files come as a set of compressed files within compressed folders. The files are available as either SAS or SPSS files. For this study, the SPSS file was downloaded. However, either option is viable as the EdSurvey R package, which was used to read in data, can handle a variety of data file types including CSV, SAS, and SPSS file types. These data were manually downloaded via a web browser onto the local computer. Then, the uncompressed data were placed into a project folder to be uncompressed before working with them. The list of data files downloaded for this study, and their respective file sizes, are shown in Appendix E.

Imported PISA Data in R

Next, the EdSurvey R package was used to build the raw data set from the now uncompressed data files. EdSurvey contains a function called readPISA() which reads the folder containing the data files and links them into a single data object which is used to start preparing the analytic sample.

Cleaned & Prepared PISA Data

The first step to preparing the analytic sample was to keyword search for potential variables of interest and select them into the analytic sample. Then, a process of data cleaning was conducted. One illustrative procedure was renaming variables to more human-readable names. For example, the variable name for a student's year of birth was changed from "st003d03t" to "birth_year". Another example was changing the name of the school-level weights from "w_schgrnrabwt" to "w_fschwt", to bring it more inline with the variable name for student-level weights (i.e., "w_fstuwt"). Another illustrative procedure was creating a variable for the ISO3 country code (e.g., "AFG" =

49

"Afghanistan") because it is easier to reference a 3-digit country code while working on the data than using the full text country name.

In addition, a number of derived variables were created. One illustrative procedure was creating a new variable called "linking_years". This variable contains calendar year values used to link that student's PISA data with the appropriately dated external country-level data. For an immigrant student, this means the linking_years variable contains a year value for their own year of immigration and the 5 years prior, inclusive. For example, an immigrant student who immigrated in the year 2010 would have linking years of 2010, 2009, 2008, 2007, 2006, and 2005. Therefore, after the data linking procedures, this same immigrant student will have added columns of external data with values for each respective linking_year (i.e., 2010 - 2005). A similar process was done for non-immigrant (i.e., native) students, but their linking_years were built from the year they took the PISA assessment (i.e., 2018) and the 5 years prior, inclusive. For example, a native student would have linking years of 2018, 2017, 2016, 2015, 2014, and 2013.

The reason for using two different 5-year windows is as follows. This 5-year window represents a time period, for both immigrant and native students, when their families either did or did not make a decision to immigrate. Therefore, these 5-year windows provided an analogous snapshot of the context of the country in which students' families were making a decision or non-decision to immigrate. Intuitively, this 5-year window makes sense for immigrant students as it encompasses a period of time right before their eventual immigration event. For native students, the intuition might be a bit harder to understand so an illustrative example is provided. For example, if a

Canadian student's family has never immigrated from Canada, as of the time of PISA 2018 data collection, then that family has continually made the decision to maintain the status quo for the past 5+ years. The ongoing decision to remain (i.e., to not immigrate) may be related to the relative stable state of the home country. However, since this study focused on only students with an immigrant background this was a moot point for natively born students.

Yet another data preparation decision was made regarding the immigrant status variable (i.e., immig_status). The immigration status that comes ready-made in PISA 2018 had some peculiarities. For example, the immigration status flag marks some students as NATIVE when those students have a different birth country and test country. Conversely, there are some students who are marked as FIRST-GENERATION immigrants but have the same birth country and test country. Since there was no additional data to clearly discern these data discrepancies, the issue was reconciled by creating a secondary immigrant status flag (i.e., immig_status2) based on either matching or not matching each participant's birth and test country. This simplified immig_status2 variable thereby provided a simpler way to identify participants as either IMMIGRANT or NATIVE. The SECOND-GENERATION category was not used in this new immigrant status variable. This left any student with matching origin and destination country labeled as NATIVE and any student with non-matching origin and destination country labeled as IMMIGRANT. Any students with a missing value for either origin or destination country (i.e., NA) were dropped from the analytic sample, since it was not possible to link external country-level data to a participant who did not have a birth or test country associated with them. Finally, this study focused on only students with an

immigrant background so the sample was further reduced to just first-generation immigrant students. In the end, the analytic sample included 9493 students with an immigrant background, 3514 schools, 42 destination countries, and 74 origin countries.

Cleaned & Prepared External Data

After creating the initial analytical sample from ILSA data, the external data were cleaned and prepared ahead of linking with the analytic sample. These procedures are described in this section.

Some of the external data was relatively ready to be worked with, requiring minimal prep work for this study (e.g., HDI, GAIN). The data prep work for these data sets included subsetting for the years of interest. Additionally, some column names were changed to make later data joining easier.

Conversely, the forced displacement data set underwent comparatively more work. The main task was stripping out metadata, pivoting the data into long-format, and joining the separate data for origin and destination counts into a single data set that contains a given country's inflow and outflow of displaced persons by country, by year. In addition, the language distance data required extensive work. The first step to creating the language distance data was to create a list of the languages used by participants in the ILSA analytic sample. This process created a list of all languages listed for participants in PISA 2018. Then, this list was filtered to remove languages that were too broad or specific to be useful for analysis or were simply not available in the ASJP database. A few illustrative examples included: "NON-HAN ETHNIC LANGUAGES (QCN)"; "OTHER FORMER YOGSLAVIAN LANGUAGES (SVN)"; "WESTERN EUROPEAN LANGUAGES"). After obtaining a list of relevant languages

from the PISA 2018 analytic sample, The ASJP language matrix was generated on a separate Windows 10 computer. A separate computer was used because the software was written with the Windows command line in mind, while the rest of this study was conducted on a Linux computer. Then this ASPJ matrix was saved and then imported into R. Next, some language name cleaning was performed. An illustrative example was changing "STANDARD_GERMAN" to just "GERMAN". Finally, language pairs and their associated language distances were extracted from the ASPJ matrix and arranged into a standard data frame object, using the PISA language list to filter for ASJP language pairs present in the analytic sample.

Joined External Data With PISA Data

After the external data were cleaned and prepared, they were joined to the ILSA data set by ISO3 country code and the appropriate linking years. An illustrative example of this linking procedure comes from the joining of the ND-GAIN data set to ILSA data set. The first country, alphabetically, in the ND-GAIN data set is Afghanistan (AFG). The preprocessed ND-GAIN has AFG data for years 2001 to 2018. Therefore, after joining, all rows for AFG participants in the ILSA data had two columns added. One column is for an ND-GAIN value for AFG as an origin country and one column for a ND-GAIN value for AFG as a destination country. The value in both new columns is the same. The purpose for having two columns with the same GAIN values, but with different designations for origin or destination country, is because a country can be either an origin country or destination country depending on a given immigrant student's individual-level immigration history.

In addition to linking the GAIN data set, the other external data sets were linked by country code and the range of linking years. This resulted in a very large, long-format data set. For example, a single participant had a row for their country-level ND-GAIN value for years 2001 to 2018, resulting in 18 rows of data per participant.

Next, some additional data processing was done. A 5-year mean value was calculated for the appropriate external country-level data for each student in the PISA data set. For immigrant students, their 5-year mean was calculated based on the 5-years prior to their year of immigration. This created a measure of a country-level variable 5 years before the immigration took place. The rationale for this decision was that the decision and actions to immigrate were likely made somewhere in that time frame. In other words, this measure provides a window into the state of the origin country at the time when decisions and actions regarding immigration were being made. For native students, their 5-year mean values were calculated based on the 5-years prior to the year of the PISA test 2018. This created a measure of a country-level variable 5 years before the PISA data collection point. The rationale for this decision to base this number on the year 2018 was to likewise create a similar 5-year window between the native and immigrant students in the analytic sample. However, in the native students case, their window goes back from their most recent year of non-immigration, compared to immigrant students whose window goes back from their most recent year of immigration. Another way to state this is that the 5-year means capture the state of the countries when native and immigrant student families were making or not making decisions to immigrate. One final process done to these 5-year means was to standardize them since they exist on very different scales.

Adjusted Weights Prior to Modeling as Suggested by PISA Documentation

An additional procedure needed to be done regarding weights. Nguyen and Kelley (2018) recommend adjusting level-1 weights before fitting mixed-effects models as unscaled weights may cause biased estimates. The EdSurvey R package uses a method from Rabe-Hesketh and Skrondal (2006) to automatically scale weights according to the respective ILSA data set; PISA in this case. However, since EdSurvey's HLM modeling function was not capable of running a 3-level model, it was not used for modeling. Therefore, this study could not rely on EdSurvey's automatic scaling of weights. Thus, weight transformation is conducted manually, using the suggested method (Nguyen & Kelley, 2018; Rabe-Hesketh & Skrondal, 2006). Scaling weights is recommended for PISA on level-1, but not for level-2 or level-3 (Nguyen & Kelley, 2018). This procedure produced an additional column in the sample data frame to hold the scaled student-level weights, keeping the original weights column as well for cross-reference. Note that country-level weights remained with a value of 1 for all countries. This was done to satisfy potential software weight input requirements while preserving the parity between each country.

Final Analytic Sample: PISA and External Data Linked

Finally, the joined analytic sample was reduced down to one row per participant, with the country-level variables representing 5-year means. There were 77 countries/territories represented in the analytic sample. This final list of countries is found in Appendix F.

Specified and Ran the Multilevel Models

Finally, the multilevel models were run. This study involved the creation of a "baseline model"/"null model"/"unconditioned model" followed by multiple models to test the association between the variables of interest and the outcome variables of PISA reading scores. Given the complexity of the data used in this study, the analytic methods require special attention. The following section on the analytic method summarizes those methods with the complete report found in Appendix G.

The Analytic Method: An Asymmetric Cross-Classified Data Structure Called For A Two-Stage Approach to Multilevel Modeling

Hierarchical PISA Data Suggests Multilevel Models

PISA data has a hierarchical structure with students nested within schools nested within countries. Data with a hierarchical structure is a candidate for multilevel modeling methods (Braun et al., 2009). Multilevel modeling encompasses a variety of very similar sub-methods including multilevel modeling, mixed modeling, or hierarchical linear modeling. Multilevel modeling affords researchers the ability to model effects at each level of the data, thus obtaining estimates of covariates and variance at each respective level (Beaton et al., 2011).

Asymmetric Cross-Classified Data Structure

To understand the methods of this study, it is important to first understand the structure of the data. With hierarchical data, such as PISA, student data can vary systematically based on how they are grouped (e.g., school of attendance; country of residence). This means it is important to pay attention to these levels in statistical modeling. There are three levels: level 1 is the student-level, level 2 is the school-level,

level 3 is the country-level. Furthermore, part of the data structure is strictly nested while another part is cross-nested. The strictly nested part is students nested within schools nested within destination countries. The cross-nested part is the students also nested within the origin countries. This asymmetric cross-classified data structure presented a challenge for statistical modeling. Figure 4 is a visualization of the data structure.

Figure 4

Structure of the Data
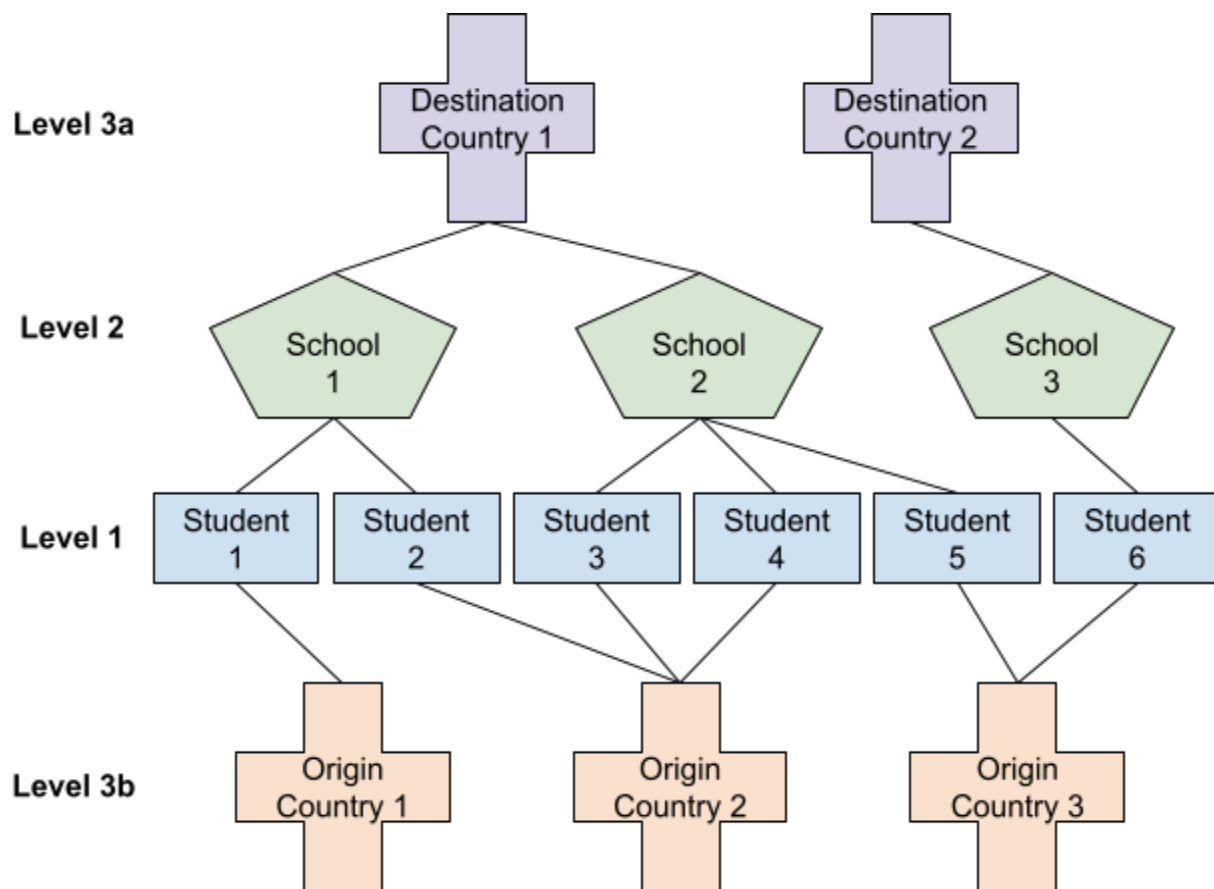
The Data Structure Presented a Challenge for Statistical Modeling

The challenge for statistical modeling was due to the available software not supporting this model structure. The desired model specification was cross-classifying level-1 within two different country-level groups (i.e., destination country; origin country). Figure 5 shows the structure of the data.

Figure 5

Structure of the Data
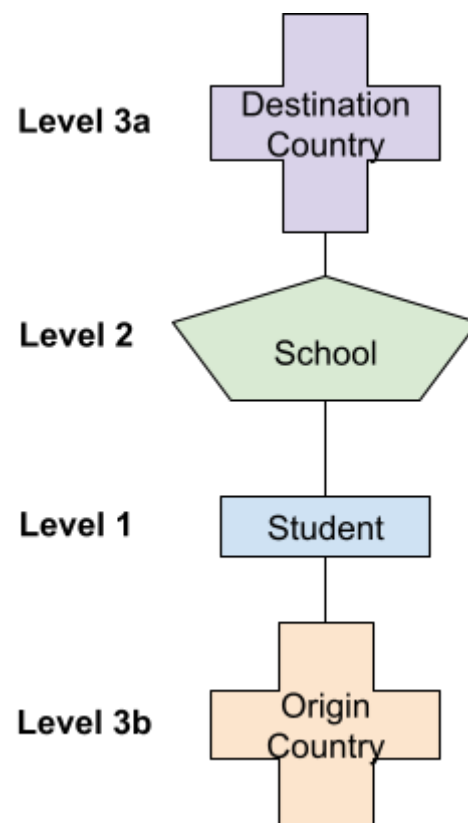
Initially, the WeMix R package was targeted for use in statistical modeling for this study (Bailey, et al., 2021). However, a limitation of the software was that it only handles fully nested data, such as the 3-level fully nested structure shown in Figure 6. But this was a problem because the data for this study has students cross-nested within two different country-levels.

Figure 6

Structure of the Data
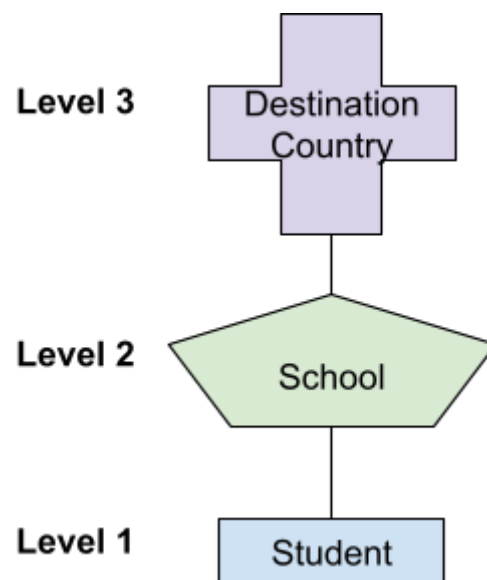
Then the SSI HLM software was targeted for statistical modeling in this study (Raudenbush & Congdon, 2021). However, while SSI HLM does support cross-classifying, it requires the cross-classifying to be level-2 cross-classified within two different level-3 groups (see Figure 7). The problem here is that level-2 (i.e., schools) isn't really nested within level-3b (i.e., origin countries). The schools are located only within destination countries, not origin countries. This modeling challenge prompted an investigation into the variance decomposition to determine whether there really is a need to model with all available levels (i.e., student, school, destination country, and origin country). The results of this investigation are explained in the section below.

Figure 7

Structure of the Data
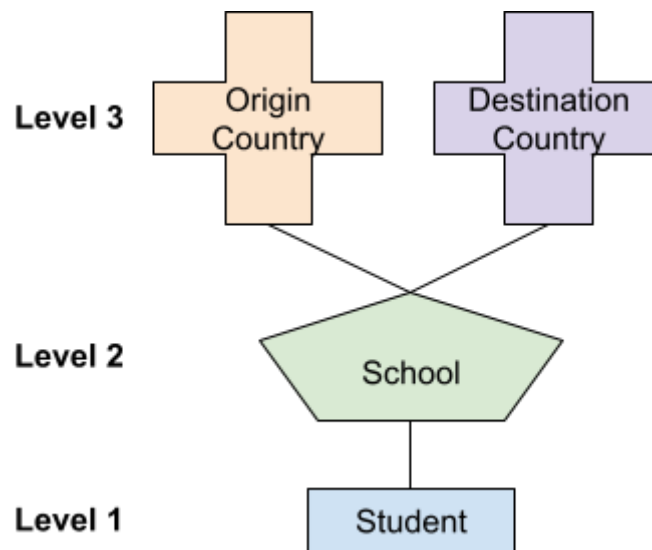
Examining Variance Proportions by Level Suggested That All Levels Are

Important

Several unconditioned models were run to investigate the variance decomposition of each level, as a way to determine the importance of each level. Both two-level and three-level models were specified for comparison. The results showed that meaningful variance was attributed to each level of data ranging from 23% to 55% of the variance within nested models. This suggested that all levels are important and should be included in the multilevel models.

The model outputs reported the proportion of variance explained by each respective level. Prior to running models, data were sorted by either destination country or origin country before modeling, producing differing results. Table 4 and Table 5 report on the first set of models (i.e., 2 levels; students nested within schools). The first column indicates the model levels. Columns 2 through 4 report the variance estimates from the WeMix R package or SSI HLM, respectively. The use of normalized weights is denoted where necessary. A run with WeMix using normalized weights was conducted to match what SSI HLM does automatically, so as to aid comparison. The results of these models indicated that school-level explained between 44% to 70% of the variance, depending on destination/origin sorting and software used. These results suggest that school-level is important for modeling immigrant students' reading achievement on PISA 2018.

Table 4

Variance Decomposition for 2-Level Unconditioned Model (Students within Schools; Destination Country Sorted)

Levels	WeMix (Non-Normalized Weights)	WeMix (Normalized Weights)	SSI HLM (Normalized Weights)
Level 1: Student	0.9388	0.5591	0.2753
Level 2: School	0.0612	0.4409	0.7250

Table 5

Variance Decomposition for 2-Level Unconditioned Model (Students within Schools; Origin Country Sorted)

Levels	WeMix (Non-Normalized Weights)	WeMix (Normalized Weights)	SSI HLM (Normalized Weights)
Level 1: Student	0.9388	0.5591	0.1923
Level 2: School	0.0612	0.4409	0.8080

Table 6 and Table 7 report on the second set of models (i.e., 2 levels; students within countries). The column descriptions are the same as the table above. The results of these models indicated that country-level explained between 23% to 24% of the variance, depending on destination/origin sorting and software used. These results suggest that both the destination and origin country-level are important for modeling immigrant students' reading achievement on PISA 2018.

Table 6

Variance Decomposition for 2-Level Unconditioned Model (Students within Countries; Destination Country Sorted)

Levels	WeMix (Non-Normalized Weights)	WeMix (Normalized Weights)	SSI HLM (Normalized Weights)
Level 1: Student	0.7663	0.7663	0.7663
Level 2: School	0.2337	0.2337	0.2337

Table 7

Variance Decomposition for 2-Level Unconditioned Model (Students within Countries; Origin Country Sorted)

Levels	WeMix (Non-Normalized Weights)	WeMix (Normalized Weights)	SSI HLM (Normalized Weights)
Level 1: Student	0.7601	0.7601	0.7601
Level 2: School	0.2399	0.2399	0.2400

Table 8 and Table 9 report on the third set of models (i.e., 3 levels; students nested within school nested within destination countries or origin countries). The column descriptions are the same as the table above. The results of these models indicated that school- and country-level combined explain between 45% to 66% of the variance. These results suggest that school-level and country-level combined are important when modeling immigrant students' reading achievement on PISA 2018.

Table 8

Variance Decomposition for 3-Level Unconditioned Model (Students within Schools with Destination Countries; Destination Country Sorted)

Levels	WeMix (Non-Normalized Weights)	WeMix (Normalized Weights)	SSI HLM (Normalized Weights)
Level 1: Student	0.7444	0.4992	0.5520
Level 2: School	0.0208	0.2747	0.2220
Level 3: Country	0.2347	0.2261	0.2265

Table 9

Variance Decomposition for 3-Level Unconditioned Model (Students within Schools within Countries; Origin Country Sorted)

Levels	WeMix (Non-Normalized Weights)	WeMix (Normalized Weights)	SSI HLM (Normalized Weights)
Level 1: Student	*NA*	*NA*	0.3858
Level 2: School	*NA*	*NA*	0.3875
Level 3: Country	*NA*	*NA*	0.2660
Note. The *NA* values are due to WeMix giving an error when trying to use the origin country as level-3: "Not a nested model; WeMix only fits nested models."			

In summary, the aforementioned structure of the data presented a modeling challenge which prompted an investigation into the variance decomposition to determine the importance of each level. Several unconditioned models were run to investigate the variance decomposition of each level, as a way to determine the importance of each level. Both two-level and three-level models were specified for comparison. The results showed that meaningful variance was attributed to each level of data ranging from 23% to 55% of the variance within 3-level models. This suggested that all levels are important and should be included in the multilevel modeling of students' reading achievement on PISA 2018. Therefore, having determined the importance of all levels, and due to the asymmetric cross-nested data structure, a two-stage approach to modeling was implemented. These stages are explained in the following sections.

Stage 1: Identifying Destination Countries with Outsized Influence; Chipping

Away Destination Country Variance

The first stage of the two-stage approach to modeling was to identify destination

countries that have an outsized impact on student achievement. This was done by

specifying a sequence of 3-level hierarchical models with students nested within

schools nested within destination countries. Figure 8 is a visualization of the model

structure.

Figure 8

Structure of the Data
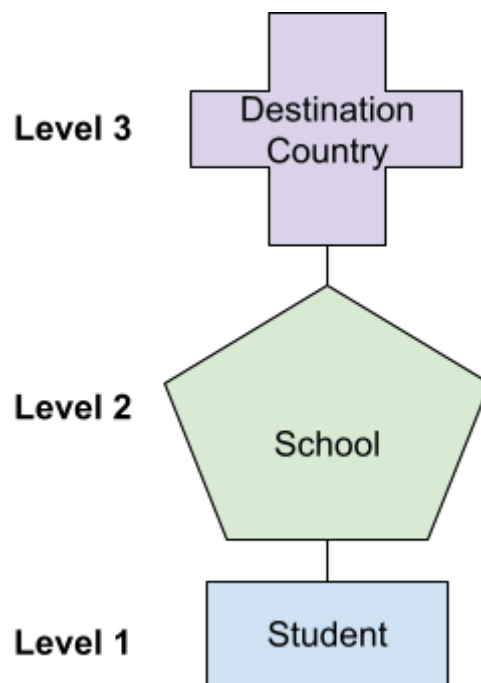
In Stage 1, three models were specified and run towards identifying destination

countries that have an outsized impact on student achievement: Model 00, Model 01,

and Model 02. For each model the following steps were taken:

1. Specify and run a 3-level hierarchical model.

2. Examine country-level residuals for outliers countries.

    a. Initial outlier threshold started with largest positive and negative residual

        values (i.e., +/-100).

3. Add outliers as fixed effects in subsequent stage 1 models.

4. Re-check variance.

5. Repeat until destination country-level variance was at 5% or less of the overall

    variance (i.e., non-significant).

Once that variance reduction goal had been achieved, twelve countries had been added

into the model as fixed effects dummy variables (i.e., Korea, Ireland, New Zealand,

Philippines, Dominican Republic, Indonesia, Canada, Turkey, Australia, Greece, North

Macedonia, and Morocco). The destination country variance reduced in each model

from 22% down to 10% and then 4%, meeting the goal of reducing the variance below

5% (see Table 10).

Table 10

Model 02 Variance Estimates Compared to Model 00

Model	Level	Variance	Variance By Level
Model 00	Student	75780	56%
	School	3004	22%
	Country	2971	22%
Model 01	Student	7575	64%
	School	3015	26%
	Country	1206	10%
Model 02	Student	7578	68%
	School	3006	27%
	Country	483	4%

The important takeaway here was this list of countries to carry forward into Stage

2, while essentially removing the destination country-level as its own level going

forward. Ultimately, the aim of this Stage 1 process was to be able to control for these

outlier countries by transforming them into fixed effects (i.e., predictor variables) at the

school-level for the models in Stage 2. The rationale for adding these countries as

level-2 variables is that each school can be reasoned as having characteristics imparted

upon it by the country the school is located within. This process solves the issue with

complex cross-nesting as it essentially removes destination country as its own level,

and instead puts the impact of destination country as a covariate within the model,

thereby allowing origin country influences to become the focal point of the modeling.

Figure 6 shows the specifications of the final Stage 1 model. The complete report of this

step-by-step Stage 1 process is found in Appendix G. Only the final Stage 1 model,

Model 02, is shown below.

Figure 6

Model 02 Specification

Summary of the model specified

Level-1 Model

$$PV1READ_{ijk} = \pi_{0jk} + e_{ijk}$$

Level-2 Model

$$\pi_{0jk} = \beta_{00k} + r_{0jk}$$

Level-3 Model

$$\beta_{00k} = \gamma_{000} + \gamma_{001}(DC\_AUS_k) + \gamma_{002}(DC\_CAN_k) + \gamma_{003}(DC\_DOM_k) + \gamma_{004}(DC\_GRC_k)$$
$$+ \gamma_{005}(DC\_IDN_k) + \gamma_{006}(DC\_IRL_k) + \gamma_{007}(DC\_KOR_k) + \gamma_{008}(DC\_MAR_k)$$
$$+ \gamma_{009}(DC\_MKD_k) + \gamma_{0010}(DC\_NZL_k) + \gamma_{0011}(DC\_PHL_k) + \gamma_{0012}(DC\_TUR_k) + u_{00k}$$

Table 11 shows the fixed effects for Model 02. The intercept (i.e., 425) represents the reference point from which to evaluate the twelve fixed effect coefficients (e.g., Australia is associated with +69 PISA reading score).

Table 11

Model 02 Fixed Effects

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
For INTRCPT2, β_{00}					
INTRCPT3, γ_{000}	424.529538	4.647458	91.347	29	<0.001
DC_AUS, γ_{001}	69.316977	22.869961	3.031	29	0.005
DC_CAN, γ_{002}	74.762934	22.787569	3.281	29	0.003
DC_DOM, γ_{003}	-142.282564	33.071557	-4.302	29	<0.001
DC_GRC, γ_{004}	-52.532902	25.269703	-2.079	29	0.047
DC_IDN, γ_{005}	-129.277461	30.553139	-4.231	29	<0.001
DC_IRL, γ_{006}	98.585864	24.190391	4.075	29	<0.001
DC_KOR, γ_{007}	103.300827	30.026244	3.440	29	0.002
DC_MAR, γ_{008}	-88.446663	26.252032	-3.369	29	0.002
DC_MKD, γ_{009}	-77.313472	33.177542	-2.330	29	0.027
DC_NZL, γ_{0010}	88.927079	23.454136	3.792	29	<0.001
DC_PHL, γ_{0011}	-124.360332	28.751305	-4.325	29	<0.001
DC_TUR, γ_{0012}	70.585712	30.886323	2.285	29	0.030

70

Table 12 shows the variance decomposition for Model 02. After adding the 12 fixed effects, the destination country variance decreased to 483.

Table 12

Model 02 Variance Components

Final estimation of level-1 and level-2 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1,r_0	54.82793	3006.10242	3472	7217.91290	<0.001
level-1, e	87.05380	7578.36408			

Final estimation of level-3 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/INTRCPT2,u_{00}	21.97516	482.90755	29	216.07663	<0.001

Stage 2: Specify and Run Cross-Classified Multilevel Models Using Variables of Interest to Explain PISA Reading Scores of Immigrant Students

The second stage of the two-stage approach to modeling was to specify a 2-level, cross-classified multilevel model. Students are cross-nested within two different level-2 groups: schools and origin countries. Figure 9 is a visualization of the data as it is to be modeled. The model is level-1 (students) cross-nested within level-2a (schools) and level-2b (origin countries). Note that the destination country is not a level itself but rather fixed effect variables at school-level. This essentially removed the destination country-level as it was controlled for by the 12 outlier countries found in Stage 1.

Figure 9

Structure of the Data
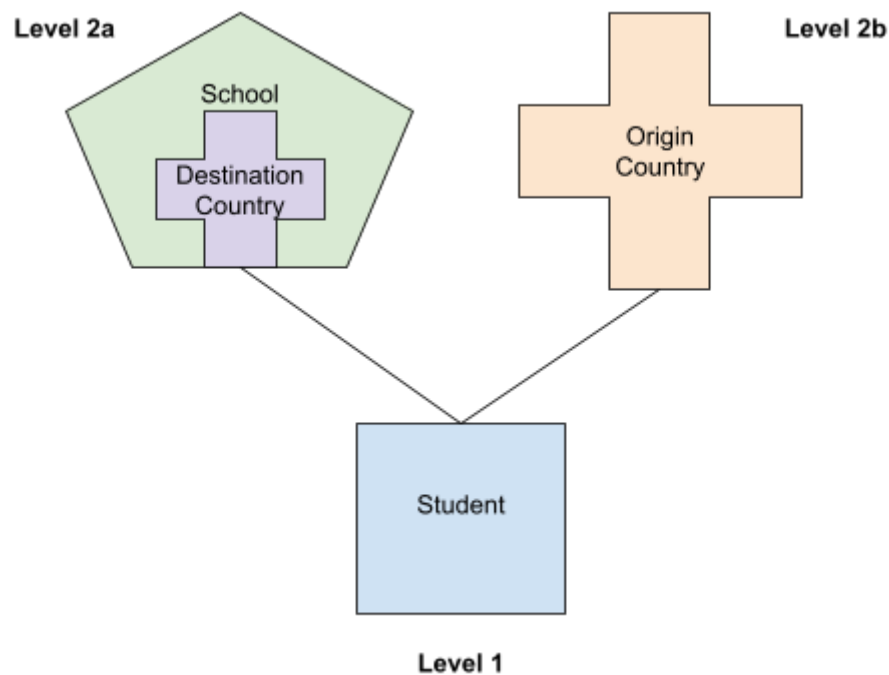
Nine total cross-nested models were specified and run for Stage 2. Each model was evaluated during the model building process. One important measure is the log-likelihood. Log-likelihood is a value output by modeling software which can be used to compare against other models. A higher value means a better model fit. Moreover, there are three commonly used goodness of fit measures based on the log-likelihood value: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Deviance. All three are used to compare models. A lower AIC, BIC, or deviance means a better model fit (Hox et al., 2017). These goodness of fit measures are used to assess whether additional predictors improve the model. A lower value means a better model. This is a way to test the significance of a variance component, by comparing the value of one model with that component compared to the same value for a model without it

(Hox et al., 2017). The collective measures of goodness of fit are used to holistically evaluate the models built for the study. It is recommended to compare the models according to a variety of criteria rather than a single fit criterion (Scott et al., 2013). Furthermore, it is informative to report how the models change from model to model rather than just report a single and final model (Scott et al., 2013).

Stage 2 started with a series of models building towards a baseline model from which to add the predictors of interest in this study. This started with the null model (i.e., Model 0). Model 1 added the 12 fixed effects countries identified in Stage 1. Model 2 added control variables (e.g., sex, socio-economic status, school socio-economic status, etc.). Model 2 was then the baseline model from which to test the variables of interest against, one at a time.

Next, the primary variables were tested one at a time in its own individual model (i.e., Model 3 = LD, Model 4 = HDI, Model 5 = GAIN, and Model 6 = FD Ratio). One of these variables was the student-level language distance while the rest were the country-level variables. Only Language Distance (i.e., Model 3) was a statistically significant coefficient.

Next, a secondary version of the country-level variables of interest were tested one at a time in its own individual model (i.e., Model 7 = Immigration Year HDI, Model 8 = Immigration Year GAIN, and Model 6 = Immigration Year FD Ratio). These were more specific versions of the country-level variables based on each student's particular year of immigration. For example, Immigration Year HDI was a measure of each student's HDI value around their unique time of immigration. This differed from the initial HDI

73

variable in that the initial version had a shared value for all students from that country. None of these models had a statistically significant coefficient.

In the end 9 total cross-nested models were specified and run for Stage 2. Table 13 compares the deviance and variance estimates between all models, which is used to assess the change in goodness of fit between models. Model 3 deviance was the lowest amongst the models. A lower deviance suggests a better model fit. A likelihood ratio test confirmed the change in deviance from the baseline model (Model 2) to Model 3 was statistically significant (p=<0.001). Additionally, the variance decreased at all levels. In the end, this modeling procedure identified just one covariate of interest that was statistically significant. This covariate was language distance from Model 3. The other variables were not statistically significant when tested individually.

Table 13

Comparison of All Models

Model	Deviance	Level	Variance	Variance By Level
Model 0 (Null)	92931	Student	3625	29%
		School	6510	51%
		Country	2518	20%
Model 1 (Null + Stage 1 Effects)	92355	Student	3638	35%
		School	5205	50%
		Country	1585	15%
Model 2 (Null + Stage 1 Effects + Controls)	91720	Student	3539	40%
		School	4293	48%
		Country	1112	12%
Model 3 (Language Distance)	**91625***	**Student**	**3488**	**39%**
		School	**4267**	**48%**
		Country	**1079**	**12%**
Model 4 (HDI)	91718	Student	3539	40%
		School	4291	48%
		Country	1089	12%

Table 13 (cont'd)

Model	Deviance	Level	Variance	Variance By Level
Model 5 (GAIN)	91716*	Student	3540	40%
		School	4289	48%
		Country	1075	12%
Model 6 (FD Ratio)	91716*	Student	3538	40%
		School	4296	48%
		Country	1041	12%
Model 7 (Immigration Year HDI)	91672*	Student	3515	33%
		School	4169	39%
		Country	3047	28%
Model 8 (Immigration Year GAIN)	91695*	Student	3506	36%
		School	4276	43%
		Country	2079	21%
Model 9 (Immigration Year FD Ratio)	91715*	Student	3528	40%
		School	4319	48%
		Country	1058	12%

Note. Asterisk after the model name denotes the model deviance is statistically significantly different from the baseline Model 2. **Bold text** denotes models with covariates that were statistically significant within the model.

The complete report of this step-by-step Stage 2 process is found in Appendix G.

Only the best fit Stage 2 model, Model 3, is shown below. Figure 10 shows the

specifications of Model 3.

Figure 10

Model 3 Specification

Level-1 Model

$$PV1READ_{ijk} = \pi_{0jk} + \pi_{1jk}*(SEX2_{ijk}) + \pi_{2jk}*(ESCS_{ijk}) + \pi_{3jk}*(LANGDIST_{ijk}) + e_{ijk}$$

Level-2 Model

$$\pi_{0jk} = \theta_0 + b_{00j} + c_{00k}$$
$$+ (\gamma_{01})*SCHSES_j$$
$$+ (\gamma_{02})*DC\_AUS_j$$
$$+ (\gamma_{03})*DC\_CAN_j$$
$$+ (\gamma_{04})*DC\_DOM_j$$
$$+ (\gamma_{05})*DC\_GRC_j$$
$$+ (\gamma_{06})*DC\_IDN_j$$
$$+ (\gamma_{07})*DC\_IRL_j$$
$$+ (\gamma_{08})*DC\_KOR_j$$
$$+ (\gamma_{09})*DC\_MAR_j$$
$$+ (\gamma_{010})*DC\_MKD_j$$
$$+ (\gamma_{011})*DC\_NZL_j$$
$$+ (\gamma_{012})*DC\_PHL_j$$
$$+ (\gamma_{013})*DC\_TUR_j$$
$$\pi_{1jk} = \theta_1$$
$$\pi_{2jk} = \theta_2$$
$$\pi_{3jk} = \theta_3$$

77

Table 14 shows the fixed effects for Model 3. The intercept (i.e., 4756 represents the reference point from which to evaluate the coefficients (e.g., language distance is associated with -0.31 PISA reading score).

Table 14

Model 3 Fixed Effects

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTERCEPT, θ_0	475.706107	10.315178	46.117	4595	<0.001
SCHSES, γ_{01}	-0.679852	0.217885	-3.120	3203	0.002
DC_AUS, γ_{02}	64.392495	13.339437	4.827	3203	<0.001
DC_CAN, γ_{03}	79.418789	13.796068	5.757	3203	<0.001
DC_DOM, γ_{04}	-122.781979	33.519623	-3.663	3203	<0.001
DC_GRC, γ_{05}	9.350883	26.667683	0.351	3203	0.726
DC_IDN, γ_{06}	-134.962667	45.652022	-2.956	3203	0.003
DC_IRL, γ_{07}	79.349199	15.663576	5.066	3203	<0.001
DC_KOR, γ_{08}	105.257943	31.457890	3.346	3203	<0.001
DC_MAR, γ_{09}	-40.201020	22.462971	-1.790	3203	0.074
DC_MKD, γ_{010}	-60.674767	26.982879	-2.249	3203	0.025
DC_NZL, γ_{011}	85.285164	16.159834	5.278	3203	<0.001
DC_PHL, γ_{012}	-83.232606	31.024108	-2.683	3203	0.007
DC_TUR, γ_{013}	0.993082	30.220247	0.033	3203	0.974
For SEX2, π_1					
INTERCEPT, θ_1	-35.802803	8.637343	-4.145	4595	<0.001
For ESCS, π_2					
INTERCEPT, θ_2	13.114632	4.630827	2.832	4595	0.005
For LANGDIST, π_3					
INTERCEPT, θ_3	-0.307951	0.134811	-2.284	4595	0.022

Table 15 shows the variance decomposition for Model 3.

Table 15

Model 3 Variance Components

Final estimation of row and level-1 variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTROW,b_{00j}	65.32497	4267.35168	3191	12762.09815	<0.001
level-1, e	59.06257	3488.38747			

Final estimation of column level variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTCOL,c_{00k}	32.84085	1078.52127	60	664.82598	<0.001

RESULTS

How Findings Answered the Research Questions

Research question #1 asked: Which specific origin country characteristics from the linked data sets have statistical significance for interpreting immigrant students' PISA reading achievement? The results reported that only one covariate from the linked data was statistically significant: Language Distance. The rest were not.

Research question #2 asked: How much additional variation in immigrant students' PISA reading achievement is explained by the linked data sets? Table 16 shows that the results reported that Model 3 explained an additional 30.2% of variance compared to the null model. Model 3 also explained an additional 1.2% of variance compared to the baseline model. Even though the proportion of additional variance explained was small (+1.2%), the applications of the language distance model were of practical significance because of the additional context that the language distance model provided around students' language use.

Table 16

Unexplained Variance Remaining By Model

Model	Variable	Unexplained Variance
M0	Null Model	12652
M1	Fixed Effects Destination Countries	10427
M2	Baseline	8944
M3	Language Distance	8834
M4	Human Development Index	8919
M5	Global Adaptation Index	8904
M6	Forced Displacement	8875
M7	Human Development Index (Immigration Year)	10731
M8	Global Adaptation Index (Immigration Year)	9861
M9	Forced Displacement (Immigration Year)	8905

First Descriptive Result: Immigration Patterns Were Asymmetric and Varied

One descriptive result was that students immigrated asymmetrically from a larger set of origin countries to a smaller set of destination countries. Figure 11 shows the distribution of countries as either: destination countries, origin countries, or both. There were 43 countries that only sent immigrant students (i.e., Origin Only). There were 32 countries that sent and received (i.e., Both). There were 10 countries that only received immigrants (i.e., Destination Only).

Figure 11

Distribution of Countries by Destination, Origin, or Both
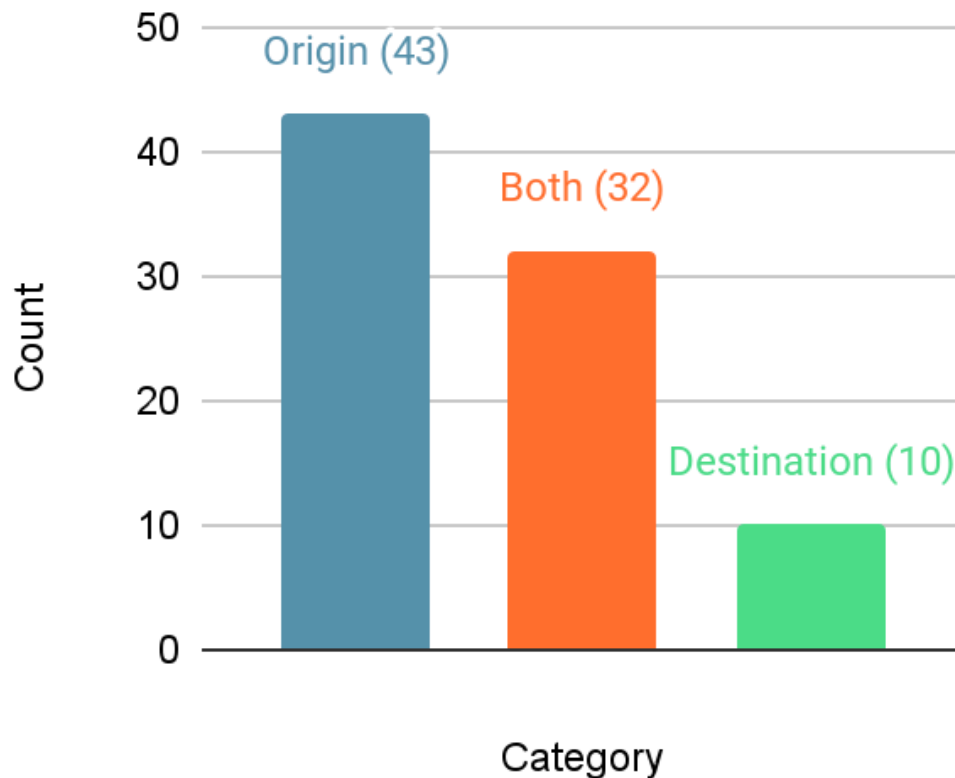
Regarding varied immigration patterns, these results show that there are different immigrant patterns for the various destination countries. For instance, the immigration movement to Canada (i.e., CAN) has little origin country overlap with Qatar (i.e., QAT). One reason for this may be due to common characteristics between origin/destination pairings. For instance, some destination countries share at least one common language with their top origin pair (e.g., Philippines to Canada). Others are physically near and share a border (e.g., Russia to Azerbaijan). Yet others may be related to multiple combined factors such as (1) a major geopolitical event, (2) sharing a land border, and (3) sharing a language (e.g., Syria to Jordan). Figure 12 is a visual representation of the

immigration patterns. Nodes placement (i.e., the boxes) were based on origin or a destination country status. For instance, China (i.e., CHN) was exclusively an origin country and thus placed far left while Australia was mostly a destination country (i.e., AUS) thus placed far right. Countries with similar immigration in and out, such as The Philippines (i.e., PHL) are placed more centrally. Node size represents case count.

Figure 12

Immigrant Paths from Destination to Origin Countries
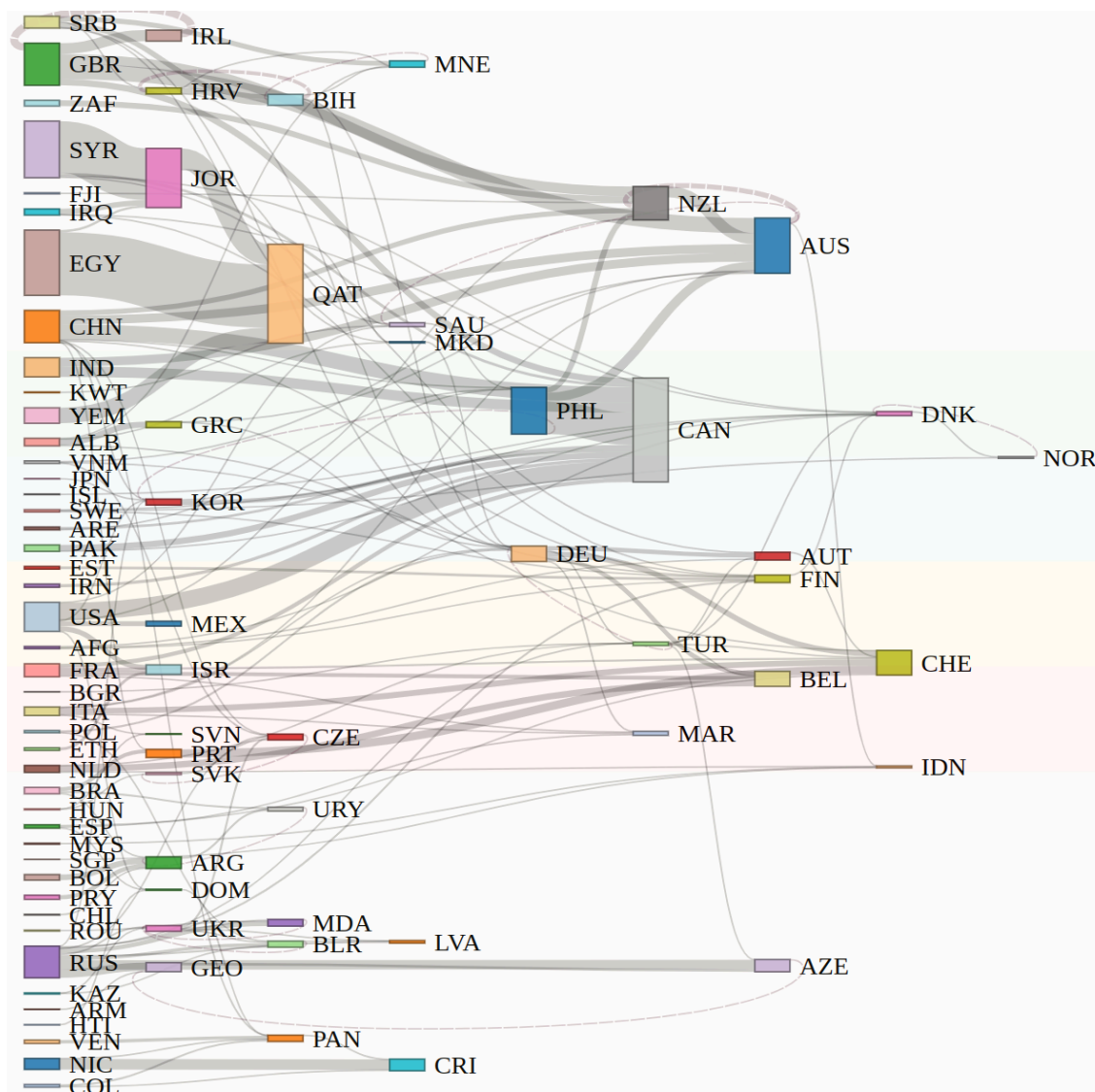
Table 17 is a tabular form of the data shown in Figure 12 above. It shows the paths of immigrant students from origin to destination country, sorted by destination country, in descending order of total immigrant count. The first two columns indicate the origin to destination country pairings. The third column denotes how many students followed that route. The final column is just a reminder for how many overall immigrants were in the sample for a given destination country. Only the Top 10 countries by overall immigrant count are shown. See Appendix H for the complete list.

Table 17

Immigrant Students Paths from Origins to Destinations, Sorted by Country

#	Destination Country	Origin Country	Origin to Dest. Count	Dest. Cntry. Immig. Count
1	CAN	PHL	526	1705
2	CAN	USA	278	1705
3	CAN	CHN	253	1705
4	CAN	IND	171	1705
5	CAN	GBR	102	1705
6	CAN	PAK	100	1705
7	CAN	KOR	73	1705
8	CAN	FRA	66	1705
9	CAN	IRN	51	1705
10	CAN	SYRA	44	1705

Table 17 (cont'd)

#	Destination Country	Origin Country	Origin to Dest. Count	Dest. Cntry. Immig. Count
11	CAN	ARE	41	1705
12	QAT	EGY	1035	1622
13	QAT	JOR	333	1622
14	QAT	YEM	254	1622
15	JOR	SYR	849	971
16	JOR	IRQ	87	971
17	JOR	EGY	35	971
18	AUS	GBR	249	905
19	AUS	NZL	177	905
20	AUS	PHL	148	905
21	AUS	CHN	143	905
22	AUS	IND	143	905
23	AUS	VNM	27	905
24	AUS	GRC	10	905
25	AUS	ITA	8	905
26	NZL	GBR	161	548
27	NZL	ZAF	93	548

Table 17 (cont'd)

#	Destination Country	Origin Country	Origin to Dest. Count	Dest. Cntry. Immig. Count
28	NZL	PHL	89	548
29	NZL	CHN	84	548
30	NZL	AUS	80	548
31	NZL	KOR	24	548
32	NZL	FJI	17	548
33	CHE	PRT	133	404
34	CHE	ITA	95	404
35	CHE	DEU	89	404
36	CHE	FRA	32	404
37	CHE	ESP	24	404
38	CHE	TUR	13	404
39	CHE	AUT	10	404
40	CHE	ALB	8	404
41	BEL	NLD	103	247
42	BEL	DEU	72	247
43	BEL	FRA	66	247
44	BEL	TUR	8	247

Table 17 (cont'd)

#	Destination Country	Origin Country	Origin to Dest. Count	Dest. Cntry. Immig. Count
45	AZE	RUS	154	201
46	AZE	GEO	42	201
47	AZE	TUR	5	201
48	ARG	BOL	91	191
49	ARG	PRY	67	191
50	ARG	BRA	15	191
51	ARG	CHL	11	191
52	ARG	URY	7	191
53	ARG	NIC	99	215
54	CRI	NIC	171	191
55	CRI	COL	14	191
56	CRI	PAN	6	191

Table 18 shows the paths of immigrant students from origin to destination country, sorted by the pairings with highest counts overall. The first two columns indicate the origin to destination country pairings. The third column denotes how many students followed that route. Only the Top 25 paths are shown, with the complete list found in Appendix Table I2. These results highlight some of the most frequent migration paths. For example, with the Top 10 migration paths, Canada appeared as an origin country three times. Another example was Jordan appearing as both an origin country and destination country.

Table 18

Immigrant Students Paths from Origins to Destinations, Sorted by Count

#	Destination Country	Origin Country	Origin to Destination Count
1	QAT	EGY	1035
2	JOR	SYR	849
3	CAN	PHL	526
4	QAT	JOR	333
5	CAN	USA	278
6	QAT	YEM	254
7	CAN	CHN	253
8	AUS	GBR	249
9	IRL	GBR	178
10	AUS	NZL	177

Table 18 (cont'd)

#	Destination Country	Origin Country	Origin to Destination Count
11	CAN	IND	171
12	CRI	NIC	171
13	NZL	GBR	161
14	AZE	RUS	154
15	AUS	PHL	148
16	AUS	CHN	143
17	AUS	IND	143
18	CHE	PRT	133
19	GEO	RUS	133
20	GBR	IRL	122
21	BEL	NLD	103
22	CAN	GBR	102
23	CAN	PAK	100
24	GRC	ALB	98
25	CHE	ITA	95

The Second Descriptive Result: Same Language Pairings Were Frequent Among Immigrant Students

A second descriptive result was the frequency of same language pairs amongst immigrant students. There were many cases of immigrant students who used the same language at home as they do in school. This suggests that language familiarity may have driven immigration choices. It may also suggest immigrant families change their home language to match the schooling language. However, the former is more likely than the later for two reasons. The main reason is supporting research suggesting that adults are unlikely to change the home language to match school language (Kang, 2013; Liu, 2018). Another reason is that the median age of immigration for students in the sample was 7 years old; likely past the age when family home language norms would have been solidified.

Table 19 which shows selected language pairing, their language distance, and the number of instances the pairings are found in the data. A minimum criteria of 20 cases or more was used for selection into this table. One aspect of this table to notice is that the most frequent language pairings are equivalent languages (n=6263). Non-equivalent language pairings are bolded within the table. Another aspect of the table to note is that some PISA countries did not capture useful data regarding the home language. For example, Canada data had 852 cases of home language recorded as "Another Language (CAN)"; the data did not provide a specific language. Therefore language distance could not be calculated for those cases.

Table 19

Language Pairs with 30+ Instances

#	Language Distance	Home Language	Test Language	Count
1	0.00	Arabic	Arabic	2245
2	0.00	English	English	1852
3	NA	Another Language (CAN)	English	852
4	0.00	Spanish	Spanish	543
5	0.00	Serbocroatian	Serbocroatian	304
6	97.76	Arabic	English	273
7	0.00	French	French	190
8	0.00	German	German	182
9	0.00	Azerbaijani	Azerbaijani	169
10	NA	Another Language (AUS)	English	162
11	102.30	Mandarin	English	154
12	0.00	Russian	Russian	125
13	NA	Another Language (NZL)	English	101

Table 19 (cont'd)

#	Language Distance	Home Language	Test Language	Count
14	0.00	Georgian	Georgian	95
15	0.00	Dutch	Dutch	78
16	NA	Another Language (CAN)	French	72
17	0.00	Hebrew	Hebrew	69
18	78.74	Portuguese	French	66
19	97.75	Albanian	Greek	60
20	97.76	English	Arabic	58
21	0.00	Portuguese	Portuguese	57
22	0.00	Romanian	Romanian	57
23	0.00	Italian	Italian	54
24	91.35	English	French	47
25	93.53	Portuguese	German	44
26	0.00	Greek	Greek	24
27	89.66	Russian	Latvian	24
28	96.06	Hindi	English	32
29	99.77	Turkish	German	32

Table 19 (cont'd)

#	Language Distance	Home Language	Test Language	Count
30	100.43	Russian	Georgian	32
31	47.55	Estonian	Finnish	31
32	0.00	Finnish	Finnish	30
33	NA	Quechua/ Guaran/ Mapuche	Spanish	30
34	96.88	English	Hebrew	28
35	95.73	German	French	27
36	100.68	Russian	Azerbaijani	27
37	0.00	Danish	Danish	26
38	96.50	Polish	German	26
39	98.83	Korean	English	26
40	100.11	Russian	Finnish	26
41	NA	Another Language (JOR)	Arabic	26
42	0.00	Korea	Korean	26
43	NA	Missing	Arabic	25

Table 19 (cont'd)

#	Language Distance	Home Language	Test Language	Count
44	0.00	Czech	Czech	24
45	61.92	Russian	Ukrainian	24
46	103.10	Vietnamese	English	23
47	55.70	Ukrainian	Czech	22
48	97.35	Cantonese	English	22
49	67.96	Portuguese	Spanish	21
50	73.23	Ethiopic	Hebrew	20
51	95.68	Albanian	German	20
52	95.88	Punjabi	English	20

Table 20 is similar to Table 19 above but now with the 0.00 language distances filtered out. When the results are filtered for just non-zero language distances, more fine grain relationships can be highlighted within the table. Illustrated examples are provided below the table.

Table 20

Language Pairs with 30+ Instances (LD=0.00 Removed)

#	Language Distance	Home Language	Test Language	Count	Countries (# of students)
1	97.76	Arabic	English	273	EGY -> QAT (203) JOR -> QAT (55) …
2	102.30	Mandarin	English	154	CHN -> AUS (75) CHN -> NZL (72) …
3	78.74	Portuguese	French	66	PRT -> CHE (66)
4	97.75	Albanian	Greek	60	ALB -> GRC (60)
5	97.76	English	Arabic	58	EGY -> QAT (32) …
6	91.35	English	French	47	USA -> CAN (20) …
7	93.53	Portuguese	German	44	PRT -> CHE (41) …
8	89.66	Russian	Latvian	34	RUS -> LVA (27) …

Table 20 (cont'd)

#	Language Distance	Home Language	Test Language	Count	Countries (# of students)
9	96.06	Hindi	English	32	IND -> AUS (31) …
10	99.77	Turkish	German	32	TUR -> AUT (11) TUR -> CHE (7) TUR -> DEU (6) …
11	100.43	Russian	Georgian	32	RUS -> GEO (31) …
12	47.55	Estonian	Finnish	31	EST -> FIN (31) …
13	96.88	English	Hebrew	28	USA -> ISR (26) …
14	95.73	German	French	27	DEU -> BEL (20) …
15	100.68	Russian	Azerbaijani	27	RUS -> AZE (23) …
16	96.50	Polish	German	26	POL -> DEU (26)
17	98.83	Korean	English	26	KOR -> NZL (22) …
18	100.11	Russian	Finnish	26	RUS -> FIN (20) …

Table 20 (cont'd)

#	Language Distance	Home Language	Test Language	Count	Countries (# of students)
19	61.92	Russian	Ukrainian	24	RUS -> UKR (19) …
20	103.10	Vietnamese	English	23	VNM -> AUS (22) …
21	55.70	Ukrainian	Czech	22	UKR -> CZE (21) …
22	97.35	Cantonese	English	22	CHN -> AUS (22)
23	67.96	Portuguese	Spanish	21	BRA -> URY (14) …
24	73.23	Ethiopic	Hebrew	20	ETH -> ISR (20)
25	95.68	Albanian	German	20	ALB -> CHE (5) ITA -> CHE (5) DEU -> CHE (4) …
26	95.88	Punjabi	English	20	IND -> AUS (16) …

Primary Inferential Result: Greater Language Distance Was Associated with Lower PISA Reading Scores

The primary inferential result of the analysis was the negative association between language distance and PISA reading scores. A one unit increase in language distance was associated with a -0.31 change in reading score, when holding all other variables constant.

To contextualize a one unit change in language distance, it is important to provide context for the language distance measure. The language distance measurement scale ranges from 0 to 104 (Wichmann et al., 2022). The larger the language distance (i.e., bigger number) between the two languages, the more dissimilar they are (Wichmann et al., 2022). A few example language pairings from Wichmann et al. (2022) are provided below for context. On the lowest end of the language distance measure are exact language pairs (e.g., home = Spanish, test = Spanish; LD = 0.00). Next, are the smallest language distances, aside from 0 values, which start at approximately 24 for this study's analytic sample. One example of very similar languages are Slovak and Czech (language distance = 33), which are highly mutually intelligible, belonging to the same Czech-Slovak language family. In the middle of the range are Portuguese and Spanish (language distance = 68) which share mutual intelligibility as they come from the same Ibero-romance language classification. An example of two dissimilar languages are Cantonese and Mandarin (language distance = 81) which are mutually unintelligible. An example of the furthest end of the language distance scale are Vietnamese and English (language distance = 103).

One way to utilize the results of the analysis, is to use the association between language distance and PISA reading scores to anticipate the expected score differential an immigrant student may encounter given their home language and the language used in the school of their destination country. Note that this score is not country-dependent as the association between PISA reading scores and language distance is modeled across all language pairs in the data set, though the model is controlled for ESCS and sex on the student-level as well as school-level ESCS.

One illustrative example of language distance in action comes from the Czech Republic. The destination country of Czech Republic received immigrant students who used Ukrainian at home (n=22). The language distance between Ukrainian and Czech is 55.70 which means that an immigrant student who used Ukranian in the home but Czech in school would be associated with a PISA reading score 17.27 points lower than immigrant students who used Czech in both locations (n=24). These values compared within their respective formulas are as follows: (Ukrainian/Czech: 55.70 * -0.31 = -17.27) vs. (Czech/Czech: 0.00 * -0.31 = 0.00).

Also within the destination country of the Czech Republic, some immigrant students used Slovak in the home (n=13). The language distance between Slovak and Czech is 32.82 which means that immigrant students who used Slovak in the home but Czech in school were associated with a PISA reading score 10.17 points lower than immigrant students who used Czech in both locations (n=24). These values compared within their respective formulas are as follows: (Slovak/Czech: 32.82 * -0.31 = -10.17) vs. (Czech/Czech: 0.00 * -0.31 = 0.00).

But the comparison between language groups is where the language distance result is most useful. Amongst immigrant students in the Czech Republic, the difference between immigrants who use Ukrainian at home and Czech in school (-17.27) versus those who use Slovak at home and Czech in school (-10.17) is 7.1 points. A 7.1 point difference corresponds to an effect size of 0.07, which can be described as small yet statistically significant effect size.

Interpreting PISA Score Differences

To interpret the significance of the point differences, such as shown above, the PISA literature provides guidance. PISA results are:

> … scaled to fit approximately normal distributions, with means around 500 score points and standard deviations around 100 score points. In statistical terms, a one-point difference on the PISA scale therefore corresponds to an effect size (Cohen's d) of 0.01; and a 10-point difference to an effect size of 0.10 (OECD, 2019a, p. 43).

Table 21 shows how to interpret a range of effect sizes.

Table 21

How to Interpret Effect Sizes Found in this Study

Cohen's *d*	Effect Size	Proportion of One Standard Deviation
0.01	Very Small	1% of a Standard Deviation
0.2	Small	20% of a Standard Deviation
0.5	Medium	50% of a Standard Deviation
0.8	Large	80% of a Standard Deviation

Table 21 (cont'd)

Cohen's *d*	Effect Size	Proportion of One Standard Deviation
1.0	Very Large	One Standard Deviation
2.0	Huge	Two Standard Deviations
Sources: (Cohen, 1988; Sawilowsky, 2009)		

The effect sizes in this study tended towards the smaller side. But effect sizes tend towards the smaller side for many educational actions. Prior research suggests the typical effect size range of 0.10 to 0.20 for the association between most education efforts and PISA scores. One study found an effect size range of -0.14 to -0.16 for the association between pre-service teacher training and PISA math scores (Carnoy et al., 2016). A second study found an effect size range of 0.10 to 0.20 for the association between student perception of instruction quality and PISA reading performance (Hu & Wang, 2022). Another study found an effect size of 0.02 for annual education reform policies (Aloisi & Tymms, 2017). Yet another study found an effect size range of -0.10 to 0.2 association between inquiry-based teaching and PISA science scores (Jerrim et al., 2022). A final study found an effect size range of 0.10 to 0.20 association between self-regulated learning and PISA reading scores (Lau & Ho, 2016).

The Language Distance Results Took On Greater Meaning Within Some Contexts Over Others

The language distance model had practical significance because of the additional context that language distance provided around students' language use. Language distance affords a closer look at the association between language and PISA reading scores in countries; especially in countries with linguistic diversity. There were some highlighted examples of the value of this additional context using four particular countries and their language pairs: Switzerland, Finland, Qatar, and Israel.

Switzerland: Multilingualism Afforded Multiple Language Distances Associations with PISA Reading Scores; Results Most Impactful for Portuguese Speakers with Effect Size of 0.09 Depending on Choice of Schooling Language

The first example context is Switzerland where Portuguese speaking immigrant students scored differently depending on their schooling language (see Table 22). The Swiss education system offers schooling in multiple languages (i.e., German, French, Italian, or Romansh) (Kużelewska, 2016). Unsurprisingly, immigrant students who used the same language in school and home did not suffer a language penalty. These students made up the reference group for interpreting the score differences. There were 121 cases of immigrants who used Portugues in the home but attended school in a different language because Portuguese speaking immigrant students didn't have a Portuguese schooling option. Amongst immigrant students who used Portuguese at home but a different language at school, using French at school (n=66) was associated with a PISA reading score -24.50 points lower than immigrant students who used the same language at both home and school (e.g., 78.74 * -0.31 = -24.50). Using Italian at

102

school (n=11) was associated with a PISA reading score -20.41 points lower than those using the same language at home and school. Using German at school (n=44) was associated with a PISA reading score -28.99 points lower.

Table 22

Score Differentials for Portuguese to French, German, & Italian

Home Language	School Language	Count	LD	Score Difference	Effect Size
French	French	38	00.00	00.00	NA
Portuguese	French	66	78.74	-24.50	-0.25
Italian	French	11	78.25	-24.26	-0.24
Italian	Italian	54	00.00	00.00	NA
Portuguese	Italian	11	65.86	-20.41	-0.20
German	German	79	00.00	00.00	NA
Portuguese	German	44	93.53	-28.99	-0.29
Albanian	German	16	95.68	-29.66	-0.30
Italian	German	12	86.30	-26.75	-0.27

While it is useful to see how each group compares to the reference group, where language distance is most useful is looking between each language group. For example, the difference between attending an Italian school and a German school was 8.58 points. This tells us that Portuguese speaking immigrant students attending an Italian language school over a German one was associated with an approximate 0.09

effect size. While these effect size differences may seem small at first glance, within the context of other studies, they are indeed impactful, since prior research suggests the typical effect size range of 0.10 to 0.20 for the association between many education efforts and PISA scores. Therefore, language distance is as impactful as many other educational actions. These results call attention to the effect that school language choice can have on immigrant students.

Finland: Language Pairings Evenly Distributed Between Finnish, Estonian, and Russian; Language Distance Between Estonian and Russian Associated with Effect Size of 0.15 on PISA Reading Scores

The second example context was Finland where Estonian and Russian speakers scored differently within Finnish schools (see Table 23). Immigrant students who used the same language in school and home (n=30) did not suffer a language penalty. These students made up the reference group for interpreting the score differences. Amongst 57 cases of immigrants who used Finnish at school but a different language at home, students using Estonian at home were associated with a PISA reading score -14.74 points lower than immigrant students who used the same language at both home and school. Students using Russian at home were associated with a PISA reading score -31.03 points lower than those using the same language at both locations. The difference between using Estonian at home versus Russian at home was 16.29 points. This tells us that speaking Russian was associated with 0.15 effect size. These results highlight the language-based disparity between Finland's two most common immigrant language groups, suggesting which groups likely require more language support.

Table 23

Score Differentials For Finnish to Estonian & Russian

Home Language	School Language	Count	LD	Score Difference	Effect Size
Finnish	Finnish	30	00.00	00.00	NA
Estonian	Finnish	31	47.55	-14.74	-0.15
Russian	Finnish	26	100.11	-31.03	-0.31

Qatar: Arabic and English Were Common Instructional Languages; English Language Instruction Associated with 0.30 effect size on PISA Reading Scores

The third example context was Qatar where 262 immigrant students chose to attend school in a different language from their home language. In Qatar, English is used as a Medium of Instruction (EMI) alongside Arabic (Hillman & Ocampo, 2018; Mustafawi & Shaaban, 2019). While most immigrant students used the same language in school and home (n=1123) and did not suffer a language penalty, there was interest in English language instruction from about 250 cases of either Egyptian or Jordanian students immigrating to Qatar to go to school in English while still using Arabic at home. However, this interest with English language schooling was associated with language-based challenges (see Table 24). Students using English at school were associated with a PISA reading score -30.31 points lower than immigrant students who used the same language at both home and school. The difference between using Arabic at home and English at school was 30.31 points. This tells us that Qatar's English language instruction option was associated with a 0.30 effect size. These results

highlight the draw that English language instruction may be having for Qatar and the anticipated effects that may follow if English instruction demand continues to grow.

Table 24

Score Differentials Based on Arabic to English & Hindi

Home Language	School Language	Count	LD	Score Difference	Effect Size
Arabic	Arabic	1233	00.00	00.00	NA
English	Arabic	51	97.76	-30.31	-0.30
Arabic	English	262	97.76	-30.31	-0.30
Hindi	Arabic	9	77.91	24.15	-0.24

Israel: Immigration from Jewish Ethiopians Led to Ethiopic and Hebrew Language Pairings; Speaking Ethiopic Languages Associated with 0.07 Effect Size Compared to English

The fourth example context was Israel. In Israel, Hebrew and Arabic are the two languages of instructions (Kelly et al., 2020). However, only Hebrew language schools were present in this data set. Immigrant students who used the same language in school and home (n=69) did not suffer a language penalty (see Table 25). These students made up the reference group for interpreting the score differences. There were 75 cases of immigrants who used Hebrew at school but a different language at home. Students using English at home were associated with a PISA reading score -30.03 points lower than immigrant students who used the same language at both home and school. Students using an Ethiopic language at home were associated with a PISA

reading score of -22.70 points lower. The difference between using English at home and Ethiopic at school was 7.33 points. This tells us that speaking Ethiopic languages was associated with an effect size of about 0.07 compared to English.

These results are meaningful because the Ethiopian immigrants in this sample are likely Jewish Ethiopians. Other research shows that Jewish Ethiopians have difficulties with post-immigration integration, resulting in economic, educational, and social stress (Berhanu, 2005; Ringel et al., 2005). However, when looking at just the language distance results, they suggested that the Jewish Ethiopian students may have a linguistic advantage over other groups, which highlights how this group could benefit less from language support and more from social support.

Table 25

Score Differentials Based on Hebrew to English, Ethiopic, & Arabic

Home Language	School Language	Count	LD	Score Difference	Effect Size
Hebrew	Hebrew	69	00.00	00.00	NA
Arabic	Hebrew	11	78.11	-24.21	-0.24
English	Hebrew	28	96.88	-30.03	-0.30
Ethiopic	Hebrew	20	73.23	-22.70	-0.23
French	Hebrew	16	93.05	-28.85	-0.29

Summary of The Results

One descriptive result was that students immigrated asymmetrically from a larger set of origin countries to a smaller set of destination countries. A second descriptive result was the frequency of same language pairs amongst immigrant students. There were many cases of immigrant students who used the same language at home as they do in school. The primary inferential result of the analysis was the negative association between language distance and PISA reading scores. A one unit increase in language distance was associated with a -0.31 change in reading score, when holding all other variables constant. The language distance model had practical significance because of the additional context that language distance provided around students' language use. Language distance affords a closer look at the association between language and PISA reading scores in countries; especially in countries with linguistic diversity. There were some highlighted examples of the value of this additional context using four particular countries and their language pairs: Switzerland, Finland, Qatar, and Israel.

DISCUSSION

The Meaning & Significance of the Principle Finding

The principal finding was the statistical significance of the language distance measure. The meaning of that finding was that linking PISA data with language distance data did indeed provide additional context for interpreting how home and school language is associated with students' PISA reading scores. This finding is significant because it affords a deeper analysis into the association between language characteristics, which were imparted by the origin and destination countries, and academic outcomes. Language distance is a continuous scale that tells us to interpret more than a simpler binary language match measure that PISA can provide. This utility was shown in the four example country contexts. This deeper analysis could not be conducted using just the PISA data alone.

The Utility of Language Distance Over Time

An affordance of language distance is its utility for measuring the variation between language groups across time. One example was shown in the Switzerland results. The Switzerland results suggested paying attention to the effect of school language choice for a particular immigrant group (i.e., Portuguese speakers). But, the groups that require attention are not static. Table 26 indicates changes in the top immigration groups within Switzerland during a recent 10-year period. First, the dominant group became less dominant (i.e., Germany). Second, two groups increased representation (i.e., Italy & France). Third, the major group of interest decreased (i.e., Portugal). Finally, a new emerging group has appeared (e.g., Romania). Together, this

highlights that as the language groups shift overtime, the utility of language distance to measure the variation between groups remains.

Table 26

Changes in Top Language Groups within Switzerland

Year: 2010-19	Year: 2020
Germany (16.8%)	Germany (14%)
Italy (10.7%)	Italy (12%)
France (9.2%)	France (11.5%)
Portugal (9.1%)	Portugal (5.5%)
Spain (4.1%)	Spain (4%)
UK (3%)	Romania (3.5%)
Poland (2.7%)	Poland (3.3%)
Source: (OECD, 2022)	

Another example of the utility of language distance across time comes from the Finland results. In Finland, these results highlighted the language-based disparity between Finland's two most common immigrant language groups (i.e., Estonian and Russian), suggesting which groups likely require more language support. The languages of instruction in Finland are primarily Finnish or, to a lesser extent, Swedish (Latomaa & Nuolijärvi, 2002). These are two languages without widespread global use which means that most immigrant students to Finland will be coming from an origin

country that does not use Finnish or Swedish. Importantly, there have been shifts in the top immigration groups over time (see Table 27). In Finland, in the year 2015, people from Estonia (16%) and Russia (10%) were the top two nationalities as a percentage of total incoming migrants. (OECD, 2017). However, by 2020, people from Estonia had decreased down to 7% while people from Russia had increased up to 15% (OECD, 2022). These trends suggest an increasing importance for Finnish educators to attend to Russian immigrants as the language distance results suggest large scoring gaps for the latter group.

Table 27

Changes in Top Language Groups within Finland

Year: 2015	Year: 2020
Estonia (16%)	Estonia (7%)
Russia (10%)	Italy (15%)
Source: (OECD, 2017; OECD, 2022)	

The Utility of Language Distance For Anticipating Language Challenges

In Qatar, the language distance results highlighted the draw that English language instruction may be having for Qatar and the anticipated effects that may follow if English instruction demand continues to grow, especially without investment in language support. The bifurcation of the Qatari educational system by either Arabic or English has caused some social discord regarding the adoption of Western, English language culture versus local Qatari traditions (Eslami et al., 2020). Furthermore,

research on trends in language use in Qatar suggest that Arabic's status is falling while the status of English has increased due to perceptions of advantages for education, business, and integration with international affairs (Ellili-Cherif & Alkhateeb, 2015). The policies that have brought English as a medium of instruction into the country have necessitated the introduction of different types of EMI approaches including foundational/bridge programs to build language skills, or bilingual programs that support the use of both languages (Eslami et al., 2020). Whichever approaches are taken, the results of this study support the usefulness of language specific academic supports for immigrant learners regardless of whether learners come from Arabic speaking homes into English speaking classrooms or vice versa.

The Utility of Language Distance for Parsing Out Language from Culture

In Israel, the language distance results identified which immigrant students likely require the most academic support, based on language characteristics of immigrating students. There were multiple language pairings between Hebrew and English, Ethiopic, French, or Arabic. Among immigrant students in Israel, the language distance associated with the smallest language deficit was Ethiopic languages. Israel received multiple periods of immigration from Jewish Ethiopians; people of Jewish ancestry residing in present-day Ethiopia (Ringel et al., 2005). Importantly, Jewish Ethiopians who immigrate to Israel have difficulties with integration into the destination country society, resulting in economic and social stress (Ringel et al., 2005). Some of the reasons for this include racial discrimination, intergenerational conflicts amongst Jewish Ethiopian families, and differences in communication styles (Ringel et al., 2005). Along educational dimensions, this particular immigrant group also lags in academic

112

performance partially due to the local school system being unfamiliar with how identity, belongingness, and negotiation of meaning take place within the cultural groups (Berhanu, 2005). In terms of language, similar results were reported in the current study. However, when looking at just the language distance results, they suggested that the Jewish Ethiopian students may have a linguistic advantage over other groups, which highlights how this group could benefit more from social support and less from language support.

The Meaning & Significance of the Analytic Method Developed for This Study

The analytic method also had meaning and significance as well. The meaning of this analytic method was that it provided one solution to the modeling challenge that the asymmetric, cross-nested data structure presented. This helped overcome the constraints of the statistical modeling software. The findings of the pre-work suggested that all levels should be addressed in the modeling of students' reading achievement on PISA 2018. The significance of this finding was that it suggested that each level of the data has something important to contribute to explaining the outcome measure. In addition, Stage 1 produced a list of 12 destination countries that have larger than typical influence on student reading scores. The significance of this finding was that these 12 destination countries could be used as fixed effects covariates, which now controlled for, solves the issue with complex cross-nesting as it essentially removes destination country as its own level, and instead puts the impact of destination country as a covariate within the model. This allowed the statistical modeling software to specify the desired models; something that was not possible prior due to the constraints of the statistical modeling software already explained in the methods section.

Connecting the Findings Back to the Literature Review

This section connects the findings of the study back to the literature review topics. The first review topic highlighted that immigrant student achievement on PISA has typically been studied using secondary analyses and that there are numerous benefits in doing so (Donnellan et al., 2011; Torney-Purta & Amadeo, 2013a). This study continued in this tradition by also conducting a secondary analysis. The second review topic was that demographic characteristics (e.g., SES, gender, language, nationality, parent occupation, etc.) were associated with PISA outcomes (Aloisi & Tymms, 2017). The findings of the study supported this prior research with the significance of a measure based on a demographic characteristic: language. The third review topic was that PISA results have shown mixed achievement amongst immigrant students (Schleicher, 2006). The findings of this study supported this prior research since the results also showed that immigrant students in this sample were a diverse subgroup with asymmetric immigration patterns, used a variety of languages, and scored differently on PISA reading. The fourth review topic was the increasing immigration in most countries (OECD, 2019a). This study did not investigate this longitudinal trend. The major critique of prior research was that secondary PISA research on immigrant students centers destination country characteristics over origin ones. This study addressed this critique by linking and then testing a set of origin measures, finding one that did indeed enrich the analysis. A minor critique of prior data linking studies was regarding the lagging adoption of studies linking origin data with education outcomes. This study addressed this research gap by entering the space and contributing a set of results towards this research area.

Implications of This Study

Focus and Design of Future Studies

First, there were implications for the focus and design of future studies. One implication is that it is important to include or at least address each level of multilevel data in related analyses. Another implication is that the multi-stage analytic method used in this study may be used to overcome software limitations of other studies with asymmetric, cross-classified data. Yet another implication is the potential to explore other data sets that can be linked with large-scale international assessments to enhance the analysis. One more implication is that efforts towards linking data at the lowest level possible may require pursuing restricted-use data which is more likely to contain identifiable information at the student-level compared to higher level data with just country-level data. While open access data sets can be easier to find and work with, there may be more benefit in obtaining access to more restricted use data, with its greater degree of identifying information and therefore more potential attachment points for linking data.

Measurement in Large-Scale Assessments

Next, are implications for measurement in large-scale assessments. The primary implication is that student-level origin data may better enhance the analysis over country-level origin data. Therefore if ILSAs ever alter their scope of data collection, student-level origin data should be increased.

Data Linking Studies

There were implications for the level at which data sets are linked. The framework presented in the review of research demonstrated that there are multiple

dimensions that serve as attachment points for which to link data (Bray & Thomas, 1995; Strietholt & Scherer, 2016). This study focused on two of those dimensions. One dimension was nonlocational demographic groupings (e.g., language distance). Another dimension was the geographic/locational dimension (e.g., country-level). The geographic/locational dimension includes seven levels ranging from the macro level to micro level: world regions/continents, countries, states/provinces, districts, schools, classrooms, and individuals. The results of this study suggest that characteristics of the student may be more influential than the origin country characteristics. For instance, the one significant measure, language distance, was a student-level measure. This interpretation aligns with the findings from the analysis of the variance decomposition of the unconditioned models which showed origin country characteristics with the lowest variance explained: student-level (~39%), school-level (~39%), and country-level (~27%). However, these same results also show that each level still merits attention for explaining immigrant students PISA reading results. Therefore, the implication is that the origin country is worth investigating and when linking PISA with origin country characteristics, those characteristics should be as close as possible to the individual-level (i.e., students) and that higher levels (i.e., countries) provide less explanatory value. The student-level language distance measure accomplished this by measuring how the origin country language carries over into home life after immigration while the particular country-level measures may have been too broad.

A counterpoint to linking at the lowest possible level is that there is a tradeoff for using lower level measure. The lower the level, the less data readily available to link to each participant in the principal data set. An illustrative example with PISA comes from

this study itself. At the most micro level of PISA (i.e., student), some of the PISA participants used very uncommon language at home (e.g., "OTHER FORMER YUGOSLAVIAN LANGUAGES") which meant the linked language distance data was missing and these students were not included in the analysis. Conversely, at the most macro level of PISA (i.e., country) every student had an associated destination and origin country which meant no students needed to be excluded for missing country-level data.

Additionally, as missing data increases, the more challenging it becomes to adequately model the outcome variable. This study dealt with this using listwise deletion, removing cases where data was missing. Therefore, when a researcher sets out to ask and answer more micro level research questions, with more selective samples, the less likely there are linkable data sets, and the less likely they will have meaningful results to report.

Practice & Policy in Education

Finally, there were implications for educational practice. One implication is to plan instruction around students' particular language pairs and corresponding language distances, instead of just origin country identity. This is in service of culturally responsive teaching (Gay, 2000; Gay, 2002). Culturally responsive teaching centers students' cultural characteristics, experiences, and perspectives (Gay, 2000; Gay, 2002). Therefore, while two students may share the same origin country identity, they can have different origin languages. Closer attention to those pairs and the degree to which they differ is a recognition of the cultural variation that exists amongst students who share a country identity. Specifically, a funds of knowledge approach to instruction is suggested.

117

The term "funds of knowledge" is in reference to the collective knowledge and skills that are historically and culturally developed by students (Moll et al., 2006). A funds of knowledge approach to instruction highlights the strengths, interests, identities, social backgrounds, and cultural backgrounds that students' bring from their origin countries (Moll et al., 2006). Teachers can incorporate linguistic funds of knowledge into their instruction. For instance, a literacy lesson can be planned around the loan words shared between students' origin and destination language. This affords immigrant students with opportunities to use their existing origin language resources while continuing to develop their origin language. Furthermore, it affirms cultural and linguistic identities which has been shown to improve academic achievement (Wu et al., 2021).

A second implication is for education policy, particularly for schools to allocate language support resources around the particular language pairs most associated with lower reading scores. Language support services encompass a wide range of services. For instance, effective language support services include many accommodations ranging from computer-administered glossaries to fully-translated assessments (Pennock-Roman & Rivera, 2011). Another example is providing translations on assessments so that language ability isn't confounded with cognitive ability (Alt et al., 2013). This can reduce cases where language learning students are misdirected towards special education services (Sanchez et al., 2010). Similar recommendations have been made for school psychologists to use cognitive assessments in multiple languages or non-verbal versions, to reduce referrals to special needs services when language services are needed (Olvera & Gomez-Cerrillo, 2011). Taking findings from this study, the results suggested that in Finland, immigrant students who use Russian at

home struggled compared to students using Estonian at home. Therefore, language accommodations such as extra time on assessments can be differentiated by particular languages (i.e., Russian speaking students require more assessment time). Another implication for education policy is regarding school accountability. It is recommended to keep a record of whether students have ever received language support services, even after exiting language support services (Robinson-Cimpian et al., 2016). This allows schools to continue to assess students' language growth past the point of language proficiency deemed acceptable by the school. Likewise, recording students' language distances provides a starting reference point from which to gauge student academic growth after they exit language support services. Retaining a record of language distances can show a school which language pairs progress faster or slower relative to each other, rather than just knowing that students have reached the schools' desired language proficiency levels. For instance, schools can know if the approximately 15 point deficit between Russian and Estonian speakers in Finland remains constant as the two groups progress through language milestones over time.

Non-Significant Findings & Reconsidering Assumptions

The non-significant findings are important to discuss in relation to both the phenomenon investigated in this study (i.e., reading achievement of students with an immigrant background on PISA) and for methodological considerations. They were also interpreted for how they prompted the researcher to reconsider prior assumptions. Prior to the analysis there was a hypothesis that PISA did not have adequate coverage of important characteristics imparted by immigrant students' origin country. Additionally

there was a hypothesis that linking origin characteristics with ILSAs would therefore enhance related analyses.

The findings of this study suggested a nuanced answer to this hypothesis. On the one hand, the language distance result did support these hypotheses because the addition of language distance allowed for a more nuanced analysis for particular countries with particular language pairs. Something that existing PISA data alone could not provide. On the other hand, the other linked measures did not enhance the analysis.

Therefore, these results prompted a reconsideration of these two assumptions. One reconsideration is whether PISA 2018 already contains sufficient depth and breadth and the PISA data overlapped whatever the linked data brought to the analysis, thereby reducing the benefit of linking additional data. For instance, perhaps PISA's measure of social, economic, and cultural status may already capture what the linked Human Development Index measure brought to the analysis. A second reconsideration is the importance of data levels for addressing PISA data collection gaps. Specifically, that any linked origin characteristics should be as close as possible to the individual-level (i.e., students) and that higher levels (i.e., countries) may provide less explanatory value; at least the ones tested within this study.

Future Research Extending from this Study

One future extension from this study is to model language distance on top of language match (i.e., same home and schooling language). This would combine the linked contribution of language distance with the already available language match measure derived from PISA data. The reason is to model the degree of contribution that language distance provides beyond a binary language match covariate. To

operationalize this, a binary language match variable would be used in the analysis. This variable captures whether each student's home and schooling languages are the same. A value of 1 is assigned to cases where the home and school language match. A value of 0 is assigned to cases where the languages do not match. Then multilevel models would be specified first using language match followed by language distance, comparing the change in model between the two.

A second future extension is to explore the non-linear effect of language distance. The reason is to explore whether the linear relationships found in this study (i.e., -0.31 coefficient for language distance) has a turning point somewhere along the slope line (e.g., effect wears off at some point). To operationalize this, a new variable would be created. This variable would be the language distance variable squared. Then multilevel models would be specified using the squared language distance variable along with the linear effect (i.e., already established language distance variable).

A third future extension is to interact the time in the destination country with language distance in a statistical model. The reason for this is to explore whether language distance matters less the longer a student is in the country. To operationalize this, a derived variable for time since immigration would need to be created. This was already done for this study during the data processing procedures.

A fourth future extension is to interpret the fixed effects for the twelve destination countries. The reason for this is to investigate what made those destinations different from the rest. To operationalize this, the countries could be interpreted in a qualitative or mixed-methods manner. As noted earlier, quantitative secondary analysis of ILSAs can

generate research questions to be answered by smaller mixed-methods studies (Torney-Purta & Amadeo, 2013a).

A fifth future extension is to switch the focus of the study from origin countries to destination countries. The reason would be to explore the characteristics imparted by the destination country which impact students' PISA reading achievement. These results could then be paired with the results from the origin country analysis to tell a more holistic story. To operationalize this, the analysis would again use a multi-stage approach to control for origin country fixed effects and explore the destination country variables that most impact student academic achievement.

One final extension of this study is to conduct a repeat analysis on the recently released PISA 2022 data. The reason would be to add a longitudinal component to this existing analysis. To operationalize this, the procedures for this study would be repeated but with improvements based on lessons learned from this initial study.

Limitations of the Study

There were a few limitations of the study to highlight. One limitation is that PISA anonymizes individual- and school-level data. This makes it harder to link data at these lower levels and easiest to link by country identity. This is important because the results of this study suggested linking data at the lowest level possible.

A second limitation was the fidelity to which immigrant students were included within the initial PISA sample. The PISA technical documents explain how the PISA sample was defined (OECD, 2019c). For instance, there were exclusion criteria for students who had insufficient experience in the assessment language (i.e., non-native speaker; limited proficiency; less than one year of instruction). However, it is uncertain

how well each country followed those criteria. Furthermore, there has been evidence

that language learners are overrepresented in special education services because

many are misidentified as such (Rueda & Windmueller, 2006; Sanatullova-Allison &

Robison-Young, 2016). This means that some immigrant students who would otherwise

be eligible for PISA could have been excluded for other reasons.

A third limitation comes from the results of sensitivity analysis. Sensitivity

analysis is a technique to examine the robustness of an inference to unobserved or

hypothetical conditions that cannot be directly addressed with the observed data (Frank

et al., 2023b). The results are useful for quantifying the terms of uncertainty in a model

due to bias from omitted variables or sampling variability (Frank et al., 2013; Frank et

al., 2023b). An online tool for sensitivity analysis was used to find how robust the

language distance result was (Rosenberg et al., 2023). To invalidate the principle

inference from Model 3, 13.79% of the estimate would have to be due to bias. This is

based on a threshold of -0.265 for statistical significance (alpha = 0.05). Therefore, to

invalidate an inference, 634 observations would have to be replaced with cases for

which the effect is 0 (RIR = 634). There is no specific threshold for interpreting the

robustness of an inference, as this is context dependent (Frank et al., 2013). However,

replacing approximately 14% of the observed cases with cases where no relationship

exists between language distance and PISA reading scores (i.e., counterfactual cases)

could invalidate the inference. This replacement rate is not so high as to be improbable.

For instance, the sample size for this study dropped from 14,246 cases to 9,493 cases

due to missing data. If these cases were not missing data, there could conceivably be at

least 634 counterfactual cases among them which could introduce sampling bias to the

degree to invalidate the inference. Generally, the higher the cost for taking action based on the inference, the higher the robustness of the inference should be (Frank et al., 2013).

A fourth limitation is related to plausible values. As already stated, the SSI HLM software does not handle plausible values in cross-nested multilevel models. Therefore, this study's compromise is to use just a single plausible value when utilizing software with PV limitations. Overall, this approach is interpreted as providing less confidence in smaller score differences (Von Davier et al., 2009).

A fifth limitation is that this study cannot make statements of causality between the explanatory variables and the outcome variable; results are limited to interpreting associations between the variables. This is because PISA is a cross-sectional design which does not control for prior achievement.

CONCLUSION

A secondary analysis of PISA 2018 data was conducted to investigate the reading achievement of students with an immigrant background. While prior research suggests the importance of demographic characteristics in secondary PISA analyses, a major critique is that PISA centers destination characteristics over origin ones, limiting the study of the association between origin characteristics and academic achievement. This study was proposed to address the critique by linking PISA data with data of origin country characteristics to allow for analyses that could not be conducted with the PISA data set alone. The study linked PISA data with data of origin country characteristics, centered the origin-based characteristics in the analyses, and then evaluated the utility of linking this additional data for explaining education outcomes. Multilevel statistical modeling was used to model the association between origin country characteristics and the academic achievement of students with an immigrant background. Linked data were: (1) Language Distance—student-level measure of similarity between home and school language; (2) Human Development Index—country-level measure of human development; (3) Global Adaptation Index—country-level measure of climate vulnerability and readiness to improve resilience/climate adaptation; and (4) forced displacement ratio—country-level ratio between inward/outward forced displacement.

The principal finding was the statistical significance of the language distance measure. This means that linking PISA data with language distance did indeed provide additional context for interpreting how home and school language is associated with students' PISA reading scores. This is significant because it affords a deeper analysis into the association between language characteristics, which were imparted by the

origin and destination countries, and students' academic outcomes. This utility was shown in four example country contexts (e.g., Switzerland, Finland, Qatar, Israel). This deeper analysis could not be conducted using just the PISA data alone. Implications were made for (1) focus and design of future studies; (2) measurement in large-scale assessments, (3) data linking studies, and (4) practice and policy in education

REFERENCES

Aloisi, C., & Tymms, P. (2017). PISA trends, social changes, and education reforms. *Educational Research and Evaluation, 23*(5-6), 180-220. https://doi.org/10.1080/13803611.2017.1455290

Alt, M., Arizmendi, G. D., Beal, C. R., & Hurtado, J. S. (2013). The effect of test translation on the performance of second grade English learners on the KeyMath‑3. *Psychology in the Schools, 50*(1), 27-36. https://doi.org/10.1002/pits.21656

Ammermueller, A. (2007). Poor background or low returns? Why immigrant students in Germany perform so poorly in the programme for international student assessment. *Education Economics, 15*(2), 215-230. https://doi.org/10.1080/09645290701263161

Andon, A., Thompson, C. G., & Becker, B. J. (2014). A quantitative synthesis of the immigrant achievement gap across OECD countries. *Large-Scale Assessments in Education, 2*(1), 1-20. https://doi.org/10.1186/s40536-014-0007-2

Arel-Bundock, V., Enevoldsen, N., & Yetman, C.J. (2018). Countrycode: an R package to convert country names and country codes. Journal of Open Source Software, 3(28), 848. https://doi.org/10.21105/joss.00848

Arikan, S., Van de Vijver, F. J., & Yagmur, K. (2017). PISA mathematics and reading performance differences of mainstream European and Turkish immigrant students. *Educational Assessment, Evaluation and Accountability, 29*(3), 229-246. https://doi.org/10.1007/s11092-017-9260-6

Azzolini, D., Schnell, P., & Palmer, J. R. (2012). Educational achievement gaps between immigrant and native students in two "new" immigration countries: Italy and Spain in comparison. *The Annals of the American Academy of Political and Social Science, 643*(1), 46-77. https://doi.org/10.1177/0002716212441590

Bailey P., Emad A., Huo H., Lee M., Liao Y., Lishinski A., Nguyen T., Xie Q., Yu J., Zhang T., Buehler, E., Bundsgaard J., C'deBaca R., & Christensen AA. (2023). EdSurvey: Analysis of NCES education survey and assessment data. R package version 3.1.0, https://www.air.org/project/nces-data-r-project-edsurvey

Bailey, P., Kelley, C., Nguyen, T., Huo, H., & Kjeldsen, C. (2021). WeMix: Weighted mixed-effects models using multilevel pseudo maximum likelihood estimation. R package version 4.0.0, https://github.com/American-Institutes-for-Research/WeMix

Bakker, D., Müller, A., Velupillai, V., Wichmann, S., Brown, C. H., Brown, P., Egorov, D., Mailhammer, R., Grant, A., & Holman, E. W. (2009). Adding typology to lexicostatistics: A combined approach to language classification. https://doi.org/10.1515/LITY.2009.009

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beaton, A.E., Rogers, A.M., Gonzalez, E., Hanly, M.B., Kolstad, A., Rust, K.F., Sikali, E., Stokes, L., and Jia, Y. (2011). *The NAEP Primer (NCES 2001-463)*. U.S. Department of Education, National Center for Education Statistics. Washington, DC.

Beenstock M., Chiswick, B.R., Repetto, G.L. (2001) The effect of linguistic distance and country of origin on immigrant language skills: application to Israel. *International Migration 39(3)*:33–60. https://doi.org/10.1111/1468-2435.00155

Berhanu, G. (2005). Normality, deviance, identity, cultural tracking and school achievement: The case of Ethiopian Jews in Israel. *Scandinavian Journal of Educational Research, 49*(1), 51-82. https://doi.org/10.1080/0031383042000302137

Braun, H., Coley, R., Jia, Y., & Trapani, C. (2009). Exploring what works in science instruction: A look at the eighth-grade science classroom. Policy Information Report. *Educational Testing Service*. https://eric.ed.gov/?id=ED507837

Bray, M., & Thomas, R. M. (1995). Levels of comparison in educational studies: Different insights from different literatures and the value of multilevel analyses. *Harvard Educational Review, 65*(3), 472-491. Retrieved from https://www.proquest.com/docview/212255856

Carnoy, M., Khavenson, T., Loyalka, P., Schmidt, W. H., & Zakharov, A. (2016). Revisiting the relationship between international assessment outcomes and educational production: evidence from a longitudinal PISA-TIMSS sample. *American Educational Research Journal, 53*(4), 1054–1085. https://doi.org/10.3102/0002831216653180

Cattaneo, M. A., & Wolter, S. C. (2015). Better migrants, better PISA results: Findings from a natural experiment. *IZA Journal of Migration, 4*(1), 1-19. https://doi.org/10.1186/s40176-015-0042-y

Chen, C., Noble, I., Hellmann, J., Coffee, J., Murillo, M., & Chawla, N. (2015). *University of Notre Dame global adaptation index. University of Notre Dame: Notre Dame, IN, USA.* https://gain.nd.edu/

Chen, C., Hellmann, J., Berrang-Ford, L., Noble, I., & Regan, P. (2018). A global assessment of adaptation investment from the perspectives of equity and efficiency. *Mitigation and Adaptation Strategies for Global Change, 23*, 101-122. https://doi.org/10.1007/s11027-016-9731-y

Chiswick, B. R., & Miller, P. W. (1994). Language choice among immigrants in a multi-lingual destination. *Journal of Population Economics, 7*(2), 119-131. https://doi.org/10.1007/BF00173615

Chiswick, B. R., & Miller, P. W. (2005). Linguistic distance: A quantitative measure of the distance between English and other languages. Journal of Multilingual and Multicultural Development, 26(1), 1-11. https://www.tandfonline.com/doi/abs/10.1080/14790710508668395

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*. Routledge. https://doi.org/10.4324/9780203771587

Corder, S. P. (1981). *Error Analysis and Interlanguage*, Oxford: Oxford University Press.

Dervis, K., & Klugman, J. (2011). Measuring human progress: the contribution of the Human Development Index and related indices. *Revue d'économie politique, 121*(1), 73-92. https://doi.org/10.3917/redp.211.0073

Donnellan, M., Trzesniewski, K. & Lucas, R. (2011) Introduction, in K. Trzewniewski, M. Donnellan & R. Lucas (Eds) *Secondary Data Analysis: an introduction for psychologists*, pp. 3-10. Washington, DC: American Psychological Association. https://doi.org/10.1037/12350-000

Dryden-Peterson, S. (2015). *The educational experiences of refugee children in countries of first asylum.* British Columbia Teachers' Federation.

Ellili-Cherif, M., & Alkhateeb, H. (2015). College students' attitude toward the medium of instruction: Arabic versus English dilemma. *Universal Journal of Educational Research Vol. 3*(3), pp. 207 - 213. DOI: 10.13189/ujer.2015.030306

Eslami, Z. R., Graham, K. M., & Bashir, H. (2020). English Medium Instruction in higher education in Qatar: A multi-dimensional analysis using the ROAD-MAPPING framework. In: *Dimova, S., Kling, J. (eds) Integrating content and language in multilingual universities, Educational Linguistics, vol 44.* 115-129. https://doi.org/10.1007/978-3-030-46947-4_7

Frank, K.A., Maroulis, S., Duong, M., & Kelcey, B. (2013). What would it take to Change an Inference?: Using rubin's causal model to interpret the robustness of causal inferences. Education, Evaluation and Policy Analysis. Vol 35: 437-460. https://doi.org/10.3102/0162373713493129

Frank, K.A., Lin, Q., Maroulis, S.J., (2023a). Embracing essential discourse in educational policy about causal inferences from observational studies: towards pragmatic social science. Handbook on Educational Policy Research. American Educational Research Association.

Frank, K.A., Lin, Q., Xu, R., Maroulis, S.J., & Mueller, A. (2023b). Quantifying the robustness of causal inferences: sensitivity analysis for pragmatic social science. *Social Science Research, 110*, 102815. https://doi.org/10.1016/j.ssresearch.2022.102815

Figueiredo, S., Alves Martins, M., & Silva, C. F. D. (2016). Second language education context and home language effect: language dissimilarities and variation in immigrant students' outcomes. International Journal of Multilingualism, 13(2). https://doi.org/10.1080/14790718.2015.1079204

Gamallo, P., Pichel, J. R., & Alegria, I. (2017). From language identification to language distance. *Physica A: Statistical Mechanics and its Applications, 484*, 152-162. https://doi.org/10.1016/j.physa.2017.05.011

Garson, G. D. (2013). *Hierarchical linear modeling: Guide and applications*. Sage.

Gay, G. (2000). *Culturally responsive teaching: Theory, research, and practice*. New York: Teachers College Press

Gay, G. (2002). Culturally responsive teaching in special education for ethnically diverse students: Setting the stage. *International Journal of Qualitative Studies in Education, 15*(6), 613-629. https://doi.org/10.1080/0951839022000014349

Ghislandi, S., Sanderson, W. C., & Scherbov, S. (2019). A simple measure of human development: The human life indicator. *Population and Development Review, 45*(1), 219. https://doi.org/10.1111/padr.12205

Giuntella, O., Kone, Z. L., Ruiz, I., & Vargas-Silva, C. (2018). Reason for immigration and immigrants' health. *Public Health, 158*, 102-109. https://doi.org/10.1016/j.puhe.2018.01.037

Goodwin, A. L. (2020). Globalization, global mindsets and teacher education. *Action in Teacher Education, 42*(1), 6-18. https://doi.org/10.1080/01626620.2019.1700848

Grimes, J. E. and Grimes, B. F. 1993. *Ethnologue: Languages of the World*, 13th edn, Dallas: Summer Institute of Linguistics, Inc.

Hart-Gonzalez, L., & Lindemann, S. (1993). Expected achievement in speaking proficiency, 1993. *School of Language Studies, Foreign Services Institute, Department of State.*

Heath, S. B., & Heath, S. B. (1983). *Ways with words: Language, life and work in communities and classrooms*. Cambridge University Press.

Hillman, S., & Ocampo Eibenschutz, E. (2018). English, super-diversity, and identity in the State of Qatar. *World Englishes, 37*(2), 228-247. https://doi.org/10.1111/weng.12312

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. Routledge.

Hu, J., & Wang, Y. (2022). Influence of students' perceptions of instruction quality on their digital reading performance in 29 OECD countries: A multilevel analysis. *Computers & Education, 189*, 104591. https://doi.org/10.1016/j.compedu.2022.104591

Jain, T. (2017). Common Tongue: The Impact of Language on Educational Outcomes. *The Journal of Economic History, 77*(2), 473-510. https://doi.org/10.1017/S0022050717000481

Jerrim, J., Oliver, M., & Sims, S. (2022). The relationship between inquiry-based teaching and students' achievement. New evidence from a longitudinal PISA study in England. *Learning and Instruction, 80*, 101310. https://doi.org/10.1016/j.learninstruc.2020.101310

Kang, H. S. (2013). Korean-immigrant parents' support of their American-born children's development and maintenance of the home language. *Early Childhood Education Journal, 41*, 431-438. https://doi.org/10.1007/s10643-012-0566-1

Kaplan, J. & Schlegel, B. (2023). fastDummies: Fast creation of dummy (binary) columns and rows from categorical variables. Version 1.7.1. https://github.com/jacobkap/fastDummies

Kelly, D.L., Centurino, V.A.S., Martin, M.O., & Mullis, I.V.S. (Eds.) (2020). TIMSS 2019 encyclopedia: Education policy and curriculum in mathematics and science. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/encyclopedia/

Klugman, J., Rodríguez, F., & Choi, H. J. (2011). The HDI 2010: new controversies, old critiques. *The Journal of Economic Inequality, 9*, 249-288. https://doi.org/10.1007/s10888-011-9178-z

Komatsu, H., & Rappleye, J. (2021). Rearticulating PISA. *Globalisation, Societies and Education, 19*(2), 245-258. https://doi.org/10.1080/14767724.2021.1878014

Kosnik, C., Beck, C., & Goodwin, A. L. (2016). Reform efforts in teacher education. In *International handbook of teacher education* (pp. 267-308). Springer, Singapore. https://doi.org/10.1007/978-981-10-0366-0_7

Kovacevic, M. (2010). Review of HDI critiques and potential improvements. *Human development research paper, 33*, 1-44.

Kużelewska, E. (2016). Language policy in Switzerland. Studies in logic, grammar and rhetoric, 45(1), 125-140. https://doi.org/10.1515/slgr-2016-0020

Latomaa, S., & Nuolijärvi, P. (2002). The language situation in Finland. *Current Issues in Language Planning, 3*(2), 95-202. https://doi.org/10.1080/14664200208668040

Lau, K. L., & Ho, E. S. C. (2016). Reading performance and self-regulated learning of Hong Kong students: What we learnt from PISA 2009. *The Asia-Pacific Education Researcher, 25*, 159-171. https://doi.org/10.1007/s40299-015-0246-1

Ledger, S., Thier, M., Bailey, L., & Pitts, C. (2019). OECD's Approach to Measuring Global Competency: Powerful Voices Shaping Education. *Teachers College Record, 121*(8), 1–40. https://doi.org/10.1177/016146811912100802

Levenshtein, V. I. (1966, February). Binary codes capable of correcting deletions, insertions, and reversals. *In Soviet physics doklady* (Vol. 10, No. 8, pp. 707-710). https://doi.org/10.1080/14664200208668040

Liu, L. (2018). "It's Just Natural": A Critical Case Study of Family Language Policy in a 1.5 Generation Chinese Immigrant Family on the West Coast of the United States. In: Sinner, M., Hult, F., Kupisch, T. (eds) Language Policy and Language Acquisition Planning. Language Policy, vol 15. Springer, Cham. p. 13-31. https://doi.org/10.1007/978-3-319-75963-0_2

Martin, M. O., Foy, P., Mullis, I. V., & O'Dwyer, L. M. (2011). Effective schools in reading, mathematics, and science at the fourth grade. *TIMSS and PIRLS*, 109-178. https://files.eric.ed.gov/fulltext/ED545256.pdf#page=117

Maskileyson, D., Semyonov, M., & Davidov, E. (2021). Economic integration of first-and second-generation immigrants in the Swiss labour market: Does the reason for immigration make a difference?. *Population, Space and Place, 27*(6), https://doi.org/10.1002/psp.2426

Moll, L., Amanti, C., Neff, D., & Gonzalez, N. (2006). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. *In Funds of knowledge* (pp. 71-87). Routledge.

Mustafawi, E., & Shaaban, K. (2019). Language policies in education in Qatar between 2003 and 2012: From local to global then back to local. *Language Policy, 18*, 209-242. https://doi.org/10.1007/s10993-018-9483-5

Nguyen, T. & Kelley, C. (2018). Methods used for estimating mixed-effects models in edsurvey. https://www.air.org/sites/default/files/EdSurvey-Mixed_Models.pdf

Nowak, A. C., & Rosenstock, T. S. (2020). *Foundations for common approaches to measure global adaptation actions in the agriculture sector: Highlights from an analysis of existing climate adaptation frameworks*. https://hdl.handle.net/10568/109718

Nilsen, T., & Gustafsson, J. E. (2014). School emphasis on academic success: exploring changes in science performance in Norway between 2007 and 2011 employing two-level SEM. *Educational Research and Evaluation, 20*(4), 308-327. https://doi.org/10.1080/13803611.2014.941371

Nyiwul, L. (2023). Climate change adaptation innovation in the water sector in Africa: Dataset. *Data in Brief*, 46, 108782. https://doi.org/10.1016/j.dib.2022.108782

OECD (2009), "Plausible Values", in *PISA Data Analysis Manual: SPSS, Second Edition*, OECD Publishing, Paris. DOI: https://doi.org/10.1787/9789264056275-7-en

OECD (2016). *PISA 2015 results, Excellence and equity in education (Vol. I)*. Paris: OECD Publishing. https://doi.org/10.1787/9789264266490-en

OECD (2017). *International migration outlook 2017*. OECD Publishing. https://doi.org/10.1787/migr_outlook-2017-en

OECD (2019a). *PISA 2018 Results (Volume I): What Students Know and Can Do*, PISA, OECD Publishing, Paris, https://doi.org/10.1787/5f07c754-en

OECD (2019b). *PISA 2018 Results (Volume II): Where All Students Can Succeed*, PISA, OECD Publishing, Paris, https://doi.org/10.1787/b5fd1b8f-en

OECD (2019c). *PISA 2018 technical report*. https://www.oecd.org/pisa/data/pisa2018technicalreport/

OECD (2020). *Global Teaching InSights: Technical Report*, OECD Publishing, Paris, http://www.oecd.org/education/school/global-teaching-insights-technical-documents.htm

OECD (2022). *International migration outlook 2022*. OECD Publishing. https://doi.org/10.1787/30fe16d2-en

Olvera, P., & Gomez-Cerrillo, L. (2011). A bilingual (English & Spanish) psychoeducational assessment MODEL grounded in Cattell-Horn Carroll (CHC)

theory: A cross battery approach. *Contemporary School Psychology: Formerly "The California School Psychologist", 15*, 117-127. https://doi.org/10.1007/BF03340968

Pennock‑Roman, M., & Rivera, C. (2011). Mean effects of test accommodations for ELLs and non‑ELLs: A meta‑analysis of experimental studies. *Educational Measurement: Issues and Practice, 30*(3), 10-28. https://doi.org/10.1111/j.1745-3992.2011.00207.x

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rabe-Hesketh, S., & Skrondal, A. (2006). Multilevel modelling of complex survey data, Journal of the Royal Statistical Society Series A: Statistics in Society, 169(4), pg. 805–827, https://doi.org/10.1111/j.1467-985X.2006.00426.x

Raudenbush, S.W., & Congdon, R.T. (2021). *HLM 8: Hierarchical linear and nonlinear modeling*. Chapel Hill, NC: Scientific Software International, Inc.

Ringel, S., Ronell, N., & Getahune, S. (2005). Factors in the integration process of adolescent immigrants: The case of Ethiopian Jews in Israel. *International Social Work, 48*(1), 63-76. https://doi.org/10.1177/0020872805048709

Robertson, S. L. (2021). Provincializing the OECD-PISA global competences project. *Globalisation, Societies and Education, 19*(2), 167-182. https://doi.org/10.1080/14767724.2021.1887725

Robinson-Cimpian, J. P., Thompson, K. D., & Umansky, I. M. (2016). Research and policy considerations for English learner equity. *Policy Insights from the Behavioral and Brain Sciences, 3*(1), 129-137. https://doi.org/10.1177/237273221562355

Rocher, T. & Hastedt, D., (2020, September). International large-scale assessments in education: a brief guide. IEA Compass: Briefs in Education. No. 10. *International Association for the Evaluation of Educational Achievement.*

Rosenberg, J. M., Narvaiz, S., Xu, R., Lin, Q., Maroulis, S., & Frank, K. A. (2023). Konfound-It!: Quantify the robustness of causal inferences (v. 2.0.0).

Rueda, R., & Windmueller, M. P. (2006). English language learners, LD, and overrepresentation: A multiple-level analysis. *Journal of Learning Disabilities, 39*(2), 99-107. https://doi.org/10.1177/00222194060390020801

Ruhose, J., & Schwerdt, G. (2016). Does early educational tracking increase migrant-native achievement gaps? Differences-in-differences evidence across

countries. *Economics of Education Review, 52*, 134-154. https://doi.org/10.1016/j.econedurev.2016.02.004

Sablan, J. R. (2019). *Can you really measure that? Combining critical race theory and quantitative methods. American Educational Research Journal, 56*(1), 178-203. https://doi.org/10.3102/0002831218798325

Sanatullova-Allison, E., & Robison-Young, V. A. (2016). Overrepresentation: An overview of the issues surrounding the identification of English language learners with learning disabilities. *International Journal of Special Education, 31*(2), n2.

Sánchez, M. T., Parker, C., Akbayin, B., & McTigue, A. (2010). Processes and Challenges in Identifying Learning Disabilities among Students Who Are English Language Learners in Three New York State Districts. Issues & Answers. REL 2010-No. 085. *Regional Educational Laboratory Northeast & Islands.*

Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied statistical methods, 8*(2), 26. https://doi.org/10.22237/jmasm/1257035100

Sharma, S. D. (2010). Making the human development index (HDI) gender-sensitive. Gender & Development, 5(1), 60-61. https://doi.org/10.1080/741922304

Schleicher, A. (2006). Where immigrant students succeed: a comparative review of performance and engagement in PISA 2003. *Intercultural Education, 17*(5), 507-516. https://doi.org/10.1080/14675980601063900

Schnepf, S. V. (2007). Immigrants' educational disadvantage: an examination across ten countries and three surveys. *Journal of population economics, 20,* 527-545. https://doi.org/10.1007/s00148-006-0102-y

Scott, M. A., Simonoff, J. S., & Marx, B. D. (2013). *The SAGE handbook of multilevel modeling*. SAGE Publications Ltd, https://doi.org/10.4135/9781446247600

Strietholt, R., & Scherer, R. (2016). The contribution of international large-scale assessments to educational research: Combining individual and institutional data sources. *Scandinavian Journal of Educational Research, 62*(3), 368-385. https://doi.org/10.1080/00313831.2016.1258729

Swadesh, Morris. (1955). Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics, 21*(2), 121-137. https://doi.org/10.1086/464321

Torney-Purta, J., & Amadeo, J. A. (2013a). International large-scale assessments: Challenges in reporting and potentials for secondary analysis. *Research in Comparative and International Education, 8*(3), 248-258. https://doi.org/10.2304/rcie.2013.8.3.248

Torney-Purta, J., & Amadeo, J. A. (2013b). The contributions of international large-scale studies in civic education and engagement. In: von Davier, M., Gonzalez, E., Kirsch, I., Yamamoto, K. (eds) *The role of international large-scale assessments: Perspectives from technology, economy, and educational research*, 87-114. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-4629-9_6

United Nations Development Programme (UNDP). (2023, May 5). Human development index. Human Development Reports. https://hdr.undp.org/data-center/human-development-index#/indicies/HDI

United Nations High Commissioner for Refugees (UNHCR). (2023, May 5). *Refugee Population Statistics Database*. Refugee Data Finder. https://www.unhcr.org/refugee-statistics/

Vargas-Montoya, L., Gimenez, G., & Fernández-Gutiérrez, M. (2023). ICT use for learning and students' outcomes: Does the country's development level matter?. *Socio-Economic Planning Sciences*, 101550. https://doi.org/10.1016/j.seps.2023.101550

Von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful. *IERI monograph Series, 2*(1), 9-36.

Werrell, C. E., Femia, F., & Sternberg, T. (2015). Did we see it coming? State fragility, climate vulnerability, and the uprisings in Syria and Egypt. *The SAIS Review of International Affairs, 35*(1), 29-46. https://www.jstor.org/stable/27000974

Wichmann, S., Holman, H.W., & Brown, C.H. (2022). The ASJP Database (version 20).

Wickham H., Averick M., Bryan J., Chang W., McGowan LD., François R., Grolemund G., Hayes A., Henry L., Hester J., Kuhn M., Pedersen TL., Miller E., Bache SM., Müller K., Ooms J., Robinson D., Seidel DP., Spinu V., Takahashi K., Vaughan D., Wilke C., Woo K., Yutani H. (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686. doi: https://joss.theoj.org/papers/10.21105/joss.01686

World Bank. (2017). Forcibly Displaced: Toward a development approach supporting refugees, the internally displaced, and their hosts: Toward a development approach supporting refugees, the internally displaced, and their hosts. Washington, DC: World Bank. http://hdl.handle.net/10986/25016

Wu, Z., Spreckelsen, T. F., & Cohen, G. L. (2021). A meta-analysis of the effect of values affirmation on academic achievement. *Journal of Social Issues, 77*(3), 702-750. https://doi.org/10.1111/josi.12415

Zhao, Y. (2018). The changing context of teaching and implications for teacher education. *Peabody Journal of Education, 93*(3), 295-308. https://doi.org/10.1080/0161956X.2018.1449896

Zhao, Y. (2020). Two decades of havoc: A synthesis of criticism against PISA. *Journal of Educational Change, 21*(2), 244-246. https://doi.org/10.1007/s10833-019-09367-x

APPENDIX A: LIST OF PROMINENT ILSAS

Table A1 shows prominent ILSAs are the organizations that administer the respective assessments.

Table A1

List of Prominent ILSAs with Administering Organization

ILSA	Organization
Global Teaching Insights (GTI)	OECD
International Civic and Citizenship Education Study (ICCS)	IEA
International Computer and Information Literacy Study (ICILS)	IEA
Programme for International Student Assessment (PISA)	OECD
Progress in International Reading Literacy Study (PIRLS)	IEA & NCES
Trends in International Mathematics and Science Study (TIMSS)	IEA

APPENDIX B: BREAKDOWN OF DEMOGRAPHIC QUESTIONS

Tables B1 to B3 show a breakdown of demographic questions by origin country or destination country for the PISA 2018 questionnaires (i.e., student, teacher, and school questionnaire). A third column is demographic questions that overlap both origin and destination.

Table B1

Breakdown of Demographic Questions on PISA 2018 Student Survey

Origin (1)	Destination (6)	Both (5)
1. ST019 In what country were you and your parents born?	1. ST011 Which of the following are in your home? 2. ST012 How many of these are there at your home? 3. ST013 How many books are there in your home? 4. ST014 The following two questions concern your mother's job: 5. ST015 The following two questions concern your father's job: 6. ST022 What language do you speak at home most of the time?	1. ST003 On what date were you born? 2. ST004 Are you female or male? 3. ST021 How old were you when you arrived in <country of test>? 4. ST023 Which language do you usually speak with the following people? 5. ST177 How many languages, including the language(s) you speak at home, do you and your parents speak well enough to converse with others?

Table B2

Breakdown of Demographic Questions on PISA 2018 Teacher Survey

Origin (0)	Destination (3)	Both (2)
1. None	1. TC186 In what country were you born? 2. TC007 How many years of work experience do you have? 3. TC014 Did you complete a teacher education or training 4. programme?	1. TC001 Are you female or male? 2. TC002 How old are you?

Table B3

Breakdown of Demographic Questions on PISA 2018 School Survey

Origin (0)	Destination (6)	Both (0)
1. None	1. SC001 Which of the following definitions best describes the 2. community in which your school is located? 3. SC002 As of <February 1, 2018>, what was the total school 4. enrolment (number of students)? 5. SC003 What is the average size of <test language> classes in 6. <national modal grade for 15-year-olds> in your 7. school? 8. SC011 Which of the following statements best describes the 9. schooling available to students in your location? 10. SC013 Is your school a public or a private school? 11. SC016 About what percentage of your total funding for a typical 12. school year comes from the following sources?	1. None

APPENDIX C: COMPLETE LIST OF VARIABLES

The variables of interest for this study included variables related to identity (e.g., student id), immigration status, language use, mathematics scores, and more. Additionally, variables for student and school weights were identified. Table C1 shows the explanatory variable, which was a student-level mathematics PISA score.

Table C1

Outcome Variable

#	PISA Variable	Description	Level	Source
1	pv1math - pv10math	The 10 plausible values for the PISA composite score of the mathematics performance subscales (e.g., algebra, score, geometry score, etc.)	Student	PISA

Table C2 shows the explanatory variables that were used for statistical modeling. These variables were mostly pre-existing, student-level variables directly from PISA with a few derived variables as well. In addition, external variables from the external data sets, mostly country-level variables with a few student-level variables, are shown here as well.

Table C2

Explanatory Variables

#	PISA Variable	Description	Level	Source
1	lang_home	Language at home	Student	PISA
2	lang_test	Language of questionnaire/assessment	Student	PISA
3	lang_match	Derived indicator of whether lang_home & lang_match are equal	Student	PISA
4	language_distance	Degree of linguistic similarity between languages in origin and destination countries.	Student	LD
5	immig_status	Immigrant status based on country of birth	Student	PISA
6	immig_status2	Derived immigrant status based on country of birth	Student	PISA
7	country_code_origin	Student country of birth (ISO3 code)	Student	PISA
8	country_code_destination	Student country of PISA test (ISO3 code)	Student	PISA

Table C2 (cont'd)

#	PISA Variable	Description	Level	Source
9	country_name _origin	Student country of birth (Full name)	Student	PISA
10	country_name _destination	Student country of PISA test (Full name)	Student	PISA
11	gain_origin_5y r_mean	Linked country-level ND-GAIN measure for student's origin country; Mean value of 5 years prior to year of immigration (immigrant) or year of test (native).	Country	GAIN
12	gain_destinati on_5yr_mean	Linked country-level ND-GAIN measure for student's destination country; Mean value of 5 years prior to year of immigration (immigrant) or year of test (native).	Country	GAIN
13	hdi_origin_5yr _mean	Linked country-level HDI measure for student's origin country; Mean value of 5 years prior to year of immigration (immigrant) or year of test (native).	Country	HDI
14	hdi_destinatio n_5yr_mean	Linked country-level HDI measure for student's destination country; Mean value of 5 years prior to year of immigration (immigrant) or year of test (native).	Country	HDI
15	refugee_ratio_ 5yr_mean_sta ndardized_orig in	Linked country-level displacement measure for student's origin country; Mean value of 5 years prior to year of immigration (immigrant) or year of test (native).	Country	WB

#	PISA Variable	Description	Level	Source
16	refugee_ratio_5yr_mean_standardized_destination	Linked country-level displacement measure for student's destination country; Mean value of 5 years prior to year of immigration (immigrant) or year of test (native).	Country	WB

Table C3 shows the control variables. These variables were a mix of student-level and school-level variables that came directly from PISA and do not include any external variables.

Table C3

Control Variables

#	PISA Variable	Description	Level	Source
1	sex	Student (Standardized) Sex	Student	PISA
2	escs	Index of economic, social and cultural status	Student	PISA
3	school_pct_ses_disadv	School-level measure of % of 15 yr-olds from socioeconomically disadvantaged homes	School	PISA
4	pct_immig_destination_country	Derived country-level measure of % of 15 yr-olds from socioeconomically disadvantaged homes	Country	PISA

Table C4 shows the weight variables that were used in statistical modeling. These weights include student-, school-, and country-level weights. The student- and school-level weights came from PISA. The adjusted student-level weight was derived from the original student-level weight. The country-level weight was added on for the purpose of running models. A value of 1 was assigned to each country since this study does not use any weighting scheme for country-level data.

Table C4

Weight Variables

#	PISA Variable	Description	Level	Source
1	W_fstuwt	Student-level weights	Student	PISA
2	w_fstuwt_adj	Student-level weights (scaled)	Student	PISA
3	w_fschwt	School-level weights	School	PISA
4	w_cntrywt	Country-level weights	Country	N/A

Table C5 shows a set of miscellaneous variables that were used for identification (e.g., student_id), used as grouping variables within modeling software (e.g., school_id), used to create derived variable (e.g., birth_year), or used for descriptive purposes (e.g., immig_count).

Table C5

Miscellaneous Variables

#	PISA Variable	Description	Level	Source
1	student_id	Student ID number	Student	PISA
2	test_year	Derived year of PISA assessment (i.e., 2018)	Student	N/A
3	birth_year	Student sex	Student	PISA
4	age	Student age	Student	PISA
5	immig_flag	Binary flag for immigration status	Student	PISA
6	immig_year	Derived year of immigration	Student	PISA
7	immig_age	Age at year of immigration	Student	PISA
8	school_id	School ID number	School	PISA
9	immig_count	Count of immigrant participants in PISA sample, per country.	Country	PISA

APPENDIX D: SOFTWARE CONSIDERED FOR STATISTICAL MODELING

Multiple software options were considered for conducting multilevel statistical modeling before selecting one. The R WeMix package (v. 4.0.0) was one software considered for conducting multilevel models (Bailey, et al., 2021). WeMix was created by the American Institutes for Research (AIR) for running HLMs with multilevel data that includes weights at multiple levels. An advantage of the WeMix package is that it can accept up to 4 levels (e.g., student, school, country, etc.). A constraint of WeMix is that it does not automatically handle plausible values (i.e., PISA has 10 PVs for mathematics). A few other software options were considered for this study on the basis of their functionality for running HLMs.

Another considered modeling software was EdSurvey, also created by AIR, which offers a function for running HLMs on multilevel data with weights at multiple levels. EdSurvey specifically focuses on some of the most well-known large-scale education surveys, such as NAEP and PISA. This means EdSurvey is specifically written for the complex sample designs, weights, and plausible values common to these types of data sets (Bailey et al., 2023). An advantage of EdSurvey is that it automatically handles plausible values and weights. A constraint of EdSurvey is that it currently only accepts two levels (e.g., student- and school-levels).

Yet another commonly used R package for multilevel models is lme4 (Bates et al, 2015). However, lme4 does not accept weights and thus is not recommended for PISA data.

A final consideration was the SSI HLM software by Scientific Software International. An advantage of the SSI HLM software is that it was developed by

prominent scholars in the multilevel modeling field. A constraint of the SSI HLM

software is that it is not a free or open source software which means it is restricted to

paying users, after a two-week trial period ends. A comparison between the considered

software is shown in Table D1.

Table D1

Modeling Software Comparison

Modeling Software	Advantages	Constraints	Used in Study
WeMix (v. 4.0.0)	- Automatically handles weights - Specify models up to 4 levels	Doesn't automatically handle PVs	Yes
EdSurvey (v. 3.1.0)	- Automatically handles weights - Automatically handles PVs	Specify models up to 2 levels	Yes
lme4 (v. 1.4-14)	None relevant for this study	Does not automatically handle weights and PVs.	No
SSI HLM (v. 8.2)	Developed by prominent scholars in multilevel modeling	Not free or open source software	Yes

APPENDIX E: LIST OF DATA FILES

Table E1

List of Downloaded PISA 2018 Data

Data File	Data File Name
Student questionnaire data file (489 MB)	SPSS_STU_QQQ.zip
School questionnaire data file (3.1 MB)	SPSS_SCH_QQQ.zip
Teacher questionnaire data file (12.8 MB)	SPSS_TCH_QQQ.zip
Cognitive item data file (466 MB)	SPSS_STU_COG.zip

APPENDIX F: FINAL LIST OF COUNTRIES USED IN STUDY

This final list of countries/territories used in this study are shown in Table F1. The first column indicates the list count. The second column contains the country/territory name. The third column shows the count for how many immigrant students were found in the countries' respective samples. The fourth column shows what percentage of the countries' samples were immigrants.

Table F1

List of Countries used in Model

#	Country	Immigrants Count in Sample	% Immigrants in Sample
1	Albania	0	0.0
2	Azerbaijan (Baku)	377	4.9
3	Argentina	215	4.7
4	Australia	1083	21.4
5	Austria	149	15.6
6	Belgium	255	13.1
7	Bosnia and Herzegovina	214	2.5
8	Brazil	0	0.0
9	Brunei Darussalam	373	6.9
10	Bulgaria	0	0.0
11	Belarus	99	3.8

Table F1 (cont'd)

#	Country	Immigrants Count in Sample	% Immigrants in Sample
12	Canada	1868	20.7
13	Chile	0	0.0
14	Chinese Taipei (Taiwan)	0	0.0
15	Colombia	0	0.0
16	Costa Rica	191	9.4
17	Croatia	80	8.8
18	Czech Republic	119	3.3
19	Denmark	118	19.4
20	Dominican Republic	82	2.4
21	Estonia	0	0.0
22	Finland	155	4.6
23	France	0	0.0
24	Georgia	150	1.0
25	Germany	82	18.7
26	Greece	110	10.3
27	Hong Kong	1007	36.8
28	Hungary	0	0.0

Table F1 (cont'd)

#	Country	Immigrants Count in Sample	% Immigrants in Sample
29	Iceland	0	0.0
30	Indonesia	37	0.2
31	Ireland	184	10.3
32	Israel	163	13.9
33	Italy	0	0.0
34	Japan	0	0.0
35	Kazakhstan	0	0.0
36	Jordan	971	17.2
37	South Korea	33	0.1
38	Lebanon	0	0.0
39	Latvia	48	4.3
40	Lithuania	0	0.0
41	Luxembourg	868	50.8
42	Macao	1074	62.4
43	Malaysia	0	0.0
44	Malta	0	0.0
45	Mexico	85	0.8

Table F1 (cont'd)

#	Country	Immigrants Count in Sample	% Immigrants in Sample
46	Moldova	112	1.2
47	Montenegro	201	5.2
48	Morocco	74	0.7
49	Netherlands	0	0.0
50	New Zealand	765	22.6
51	Norway	34	7.0
52	Panama	165	4.7
53	Peru	0	0.0
54	Philippines	42	0.6
55	Poland	0	0.0
56	Portugal	67	4.8
57	Qatar	1711	40.3
58	Romania	0	0.0
59	Russian Federation	0	0.0
60	Saudi Arabia	66	7.4
61	Serbia	0	0.0
62	Singapore	0	0.0

Table F1 (cont'd)

#	Country	Immigrants Count in Sample	% Immigrants in Sample
63	Slovak Republic	46	0.8
64	Vietnam	0	0.0
65	Slovenia	15	4.4
66	Spain	0	0.0
67	Sweden	0	0.0
68	Switzerland	443	30.6
69	Thailand	0	0.0
70	United Arab Emirates	0	0.0
71	Turkey	26	0.6
72	Ukraine	43	1,9
73	North Macedonia	24	1.1
74	United Kingdom	149	8.0
75	United States	0	0.0
76	Uruguay	63	1.0
77	B-S-J-Z (CHINA)	0	0.0

APPENDIX G: COMPLETE STEPS FOR ANALYTICAL METHOD

This study involves the creation of a baseline model called the "null model" or "unconditioned model" followed by multiple models to test the association between the variables of interest and the outcome variables of PISA reading scores. This appendix contains the complete methods already summarized in the main body text.

Stage 1: Identifying Destination Countries with Outsized Influence; Chipping Away Destination Country Variance

Model 00: Null Model

First, a null model was specified and run. Double digit numbering convention was used to differentiate between models in Stage 1 and Stage 2. Figure G1 shows the model specification.

Figure G1

Model 00 Specification

Summary of the model specified

Level-1 Model

$$PV1READ_{ijk} = \pi_{0jk} + e_{ijk}$$

Level-2 Model

$$\pi_{0jk} = \beta_{00k} + r_{0jk}$$

Level-3 Model

$$\beta_{00k} = \gamma_{000} + u_{00k}$$

Mixed Model

$$PV1READ_{ijk} = \gamma_{000} + r_{0jk} + u_{00k} + e_{ijk}$$

Table G2 shows the fixed effects for Model 00. In the fixed effects, there are no covariates in the null model so the only relevant term is the intercept (i.e., 423) which represents the grand mean reading score for all students in the data.

Table G2

Model 00 Fixed Effects

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
For INTRCPT2, β_{00}					
INTRCPT3, γ_{000}	423.365309	8.698729	48.670	41	<0.001

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
For INTRCPT2, β_{00}					
INTRCPT3, γ_{000}	423.365309	8.697765	48.675	41	<0.001

Table G3 shows the variance decomposition for the null model. Most of the variance decomposition belongs to level-1 (i.e. 7580), with an equal amount of variance at both level-2 (i.e., 3004) and level-3 (i.e., 2971).

Table G3

Model 00 Variance Components

Final estimation of level-1 and level-2 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1,r_0	54.81244	3004.40307	3472	7216.55552	<0.001
level-1, e	87.06179	7579.75501			

Final estimation of level-3 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/INTRCPT2,u_{00}	54.50516	2970.81220	41	1569.41415	<0.001

Table G4 shows the model variance decomposition. The first column indicates the model name. The second column indicates model levels. The third column indicates the variance by level. The fourth column indicates the variance decomposition by level.

Table G4

Model Variance Decomposition

Model	Level	Variance	Variance By Level
Model 00	Student	75780	56%
	School	3004	22%
	Country	2971	22%

158

Examining Residuals For Outliers

The residuals of the aforementioned Model 00 were examined. Particular attention was given to level-3 residuals. The reason for this is to identify destination countries that may be significantly different from the rest. These countries would be added into subsequent Stage 1 models (e.g., Model 01, Model 02) as fixed effects dummy variables. The goal was to reduce the variance for the destination countries until it becomes small (i.e., non-significant) at less than 5% of the overall variance. Figure G2 is a plot of the level 3 residuals versus the fitted values. The x-axis is the fitted values scale while the y-axis is the residuals scale. Here, the expectation is for the residuals to be scattered randomly around the horizontal zero line at y = 0. A random scattering of points suggests that the residuals have no systematic relationship with the fitted values, meeting the assumption of linearity necessary for linear modeling. Figure G2 most obviously indicates the categorical nature of the level 3 residuals as they are plotted with respect to the rest of the destination countries which all share a mean effect in the null model. In other words, these residuals represent how much each destination country deviated from the predicted value that was itself based on the mean of all the countries. The figure indicated a large difference in deviation from the destination country mean residual. However, the points are generally distributed symmetrically around the zero line. Therefore, this figure supports the assumption of linearity for level 3 residuals.

Figure G2

Residuals vs. Fitted Plot for Level 3 (Destination Countries)
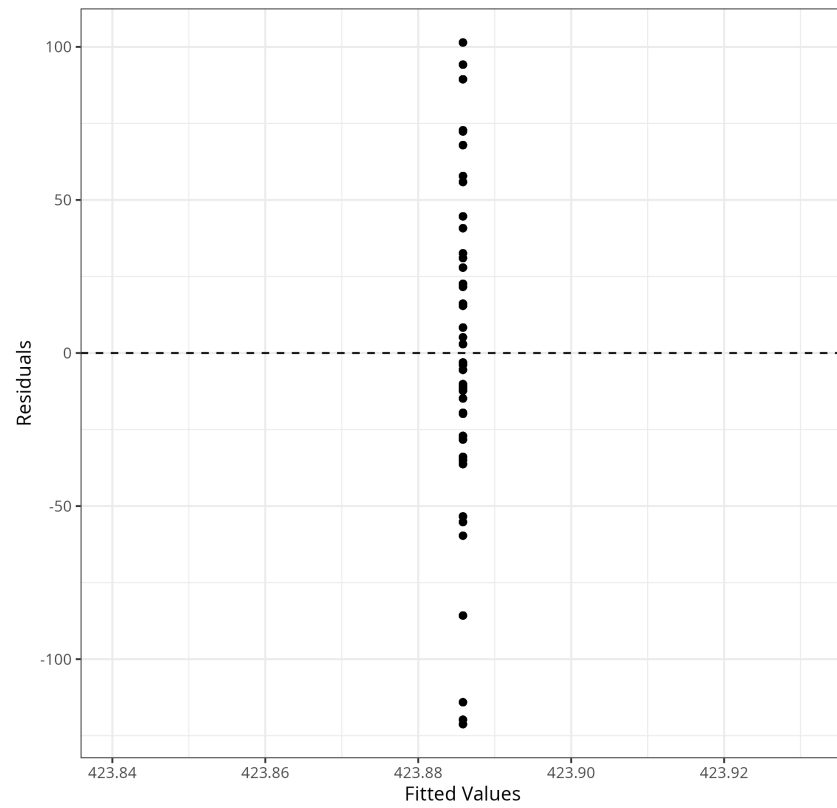
Figure G3 is a normal Q-Q plot of the level 3 residuals. Some of the residual

points fall near the reference line. However, the tails of the residual data are heavy and

deviate from the reference line. Therefore, this indicates non-normality of the level 3

residuals due to outliers.

Figure G3

Normal Q-Q Plot of Residuals for Level 3 (Destination Countries)
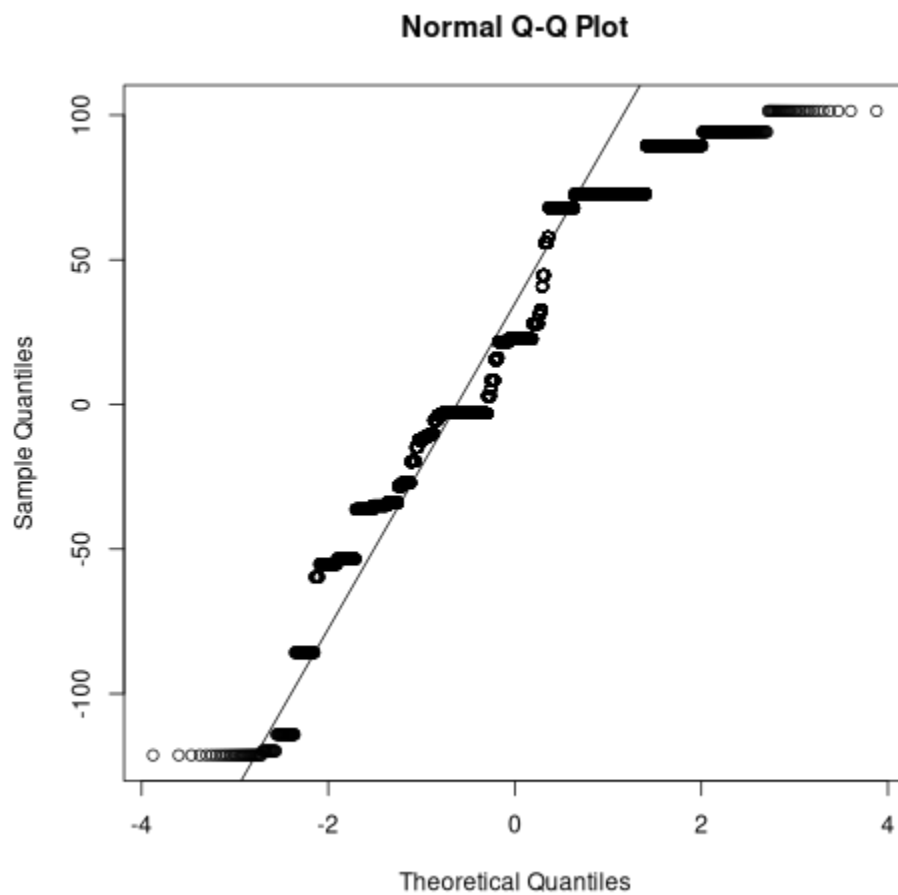
Figure G4 is a histogram of the level 3 residuals. Figure G4 also suggests non-normality (e.g., bimodal) for the level 3 residuals, due to negative residual outliers on the left of the figure and the positive frequency outliers to the right.

Figure G4

Histogram of Residuals for Level 3 (Destination Countries)
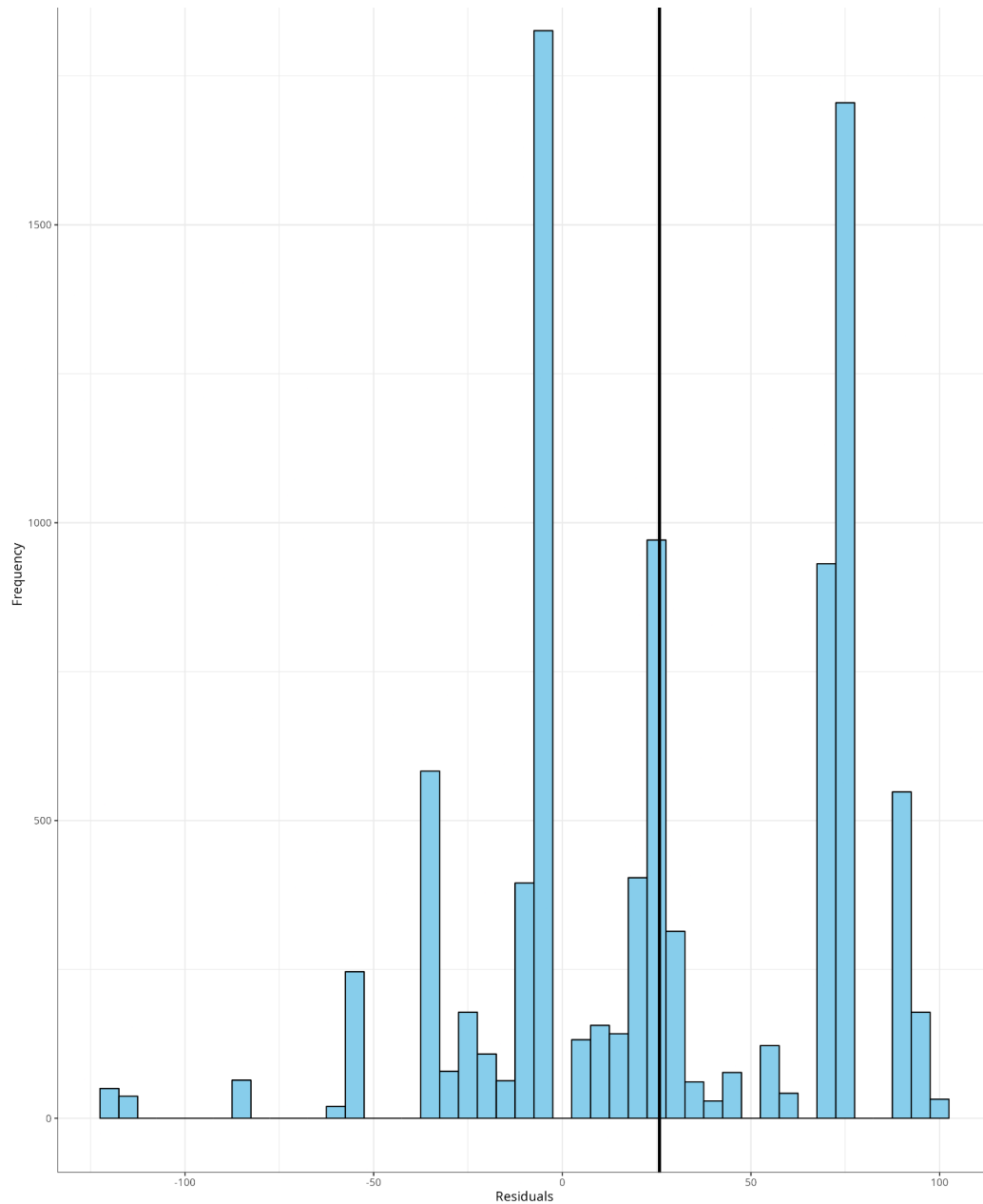
The level-3 residuals are shown below in Table G5. The first column shows the destination country name while the second column shows the value of the residual for that country. Rows are sorted from highest residual value to lowest.

Table G5

Destination Country Residual Values

#	Country	Residual Value
1	Korea	+102
2	Ireland	+94
3	New Zealand	+89
4	Canada	+73
5	Turkey	+72
6	Australia	+68
7	Latvia	+58
8	Great Britain	+56
9	Croatia	+45
10	Norway	+41
11	Portugal	+33
12	Denmark	+31
13	Belgium	+28
14	Jordan	+23

Table G5 (cont'd)

#	Country	Residual Value
15	Switzerland	+22
16	Ukraine	+16
17	Belarus	+15
18	Israel	+8
19	Slovenia	+5
20	Finland	+3
21	Qatar	-3
22	Czech Republic	-4
23	Moldova	-5
24	Montenegro	-10
25	Panama	-11
26	Saudi Arabia	-11
27	Austria	-12
28	Uruguay	-15
29	Germany	-20
30	Slovakia	-20
31	Bosnia & Herzegovina	-27

Table G5 (cont'd)

#	Country	Residual Value
32	Mexico	-28
33	Argentina	-34
34	Azerbaijan	-35
35	Costa Rica	-36
36	Georgia	-53
37	Greece	-55
38	North Macedonia	-60
39	Morocco	-86
40	Philippines	-114
41	Dominican Republic	-120
42	Indonesia	-121

Chipping Away Destination Country Variance

The purpose of stage 1 was to examine the residuals to identify which destination countries might differ from the rest. Ultimately, the aim of this Stage 1 process was to be able to control for these outlier countries by transforming them into fixed effects (i.e., predictor variables) at the school-level for the models in Stage 2. The rationale for adding these countries as level-2 variables is that each school can be reasoned as having characteristics imparted upon it by the country the school is located within. This

165

process solves the issue with complex cross-nesting as it essentially removes destination country as its own level, and instead puts the impact of destination country as a covariate within the model, thereby allowing origin country influences to become the focal point of the modeling.

Model 01: Null Model + Destination Country Effects

Model 01 was specified and run. Figure G5 shows the model specification. Six destination countries were added as fixed effects to the destination country-level. These 6 are the destination countries identified in Model 00 as having the largest residual absolute values (i.e., Korea, Ireland, New Zealand, Philippines, Dominican Republic, and Indonesia). They were entered into Model 01 as binary dummy variables at the country-level. For example, the variable for Korea takes on a value of either 1 (i.e., "Yes") if the country is Korea or a value of 0 (i.e., "No") if it is not.

Figure G5

Model 01 Specification

Summary of the model specified

Level-1 Model

$$PV1READ_{ijk} = \pi_{0jk} + e_{ijk}$$

Level-2 Model

$$\pi_{0jk} = \beta_{00k} + r_{0jk}$$

Level-3 Model

$$\beta_{00k} = \gamma_{000} + \gamma_{001}(DC\_DOM_k) + \gamma_{002}(DC\_IDN_k) + \gamma_{003}(DC\_IRL_k) + \gamma_{004}(DC\_KOR_k)$$
$$+ \gamma_{005}(DC\_NZL_k) + \gamma_{006}(DC\_PHL_k) + u_{00k}$$

Table G7 shows the fixed effects for Model 01. The intercept (i.e., 425) represents the reference point from which to evaluate the six coefficients (e.g., Korea is associated with +102 PISA reading score). This means that immigrant students who settled in Korea were associated with scores 102 points higher than the mean score of students attending school in any other country (i.e., 425); except for those six countries modeled as fixed effects.

Table G7

Model 01 Fixed Effects

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
For INTRCPT2, β_{00}					
INTRCPT3, γ_{000}	425.470542	6.218289	68.422	35	<0.001
DC_DOM, γ_{001}	-143.212284	42.829752	-3.344	35	0.002
DC_IDN, γ_{002}	-130.220232	40.900677	-3.184	35	0.003
DC_IRL, γ_{003}	97.640889	36.396685	2.683	35	0.011
DC_KOR, γ_{004}	102.356272	40.512490	2.527	35	0.016
DC_NZL, γ_{005}	87.982156	35.918466	2.449	35	0.019
DC_PHL, γ_{006}	-125.302822	39.576615	-3.166	35	0.003

Table G8 shows the variance decomposition for Model 01. After adding the 6

fixed effects, the destination country variance decreased to 1206.

Table G8

Model 01 Variance Components

Final estimation of level-1 and level-2 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1,r_0	54.91108	3015.22668	3472	7221.34920	<0.001
level-1, e	87.03381	7574.88465			

Final estimation of level-3 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/INTRCPT2,u_{00}	34.72104	1205.55062	35	1028.10531	<0.001

After adding the destination country fixed effects to Model 00, country-level

variance value decreased from 22% to 10%, not yet under the target 5% or less of the

overall variance (see Table G9).

Table G9

Model Variance Decomposition

Model	Level	Variance	Variance By Level
Model 00	Student	75780	56%
	School	3004	22%
	Country	2971	22%

Table G9 (cont'd)

Model	Level	Variance	Variance By Level
Model 01	Student	7575	64%
	School	3015	26%
	Country	1206	10%

Model 02: Null Model + More Destination Country Effects

Model 02 was specified and run. Figure G6 shows the model specification. Six

additional countries were added as fixed effects to the destination country-level. They

were the destination countries identified in Model 00 with next largest residual absolute

values (i.e., Canada, Turkey, Australia, Greece, North Macedonia, and Morocco). They

were entered into Model 02 as binary dummy variables at the country-level.

Figure G6

Model 02 Specification

Summary of the model specified

Level-1 Model

$$PV1READ_{ijk} = \pi_{0jk} + e_{ijk}$$

Level-2 Model

$$\pi_{0jk} = \beta_{00k} + r_{0jk}$$

Level-3 Model

$$\begin{aligned}\beta_{00k} = {} & \gamma_{000} + \gamma_{001}(DC\_AUS_k) + \gamma_{002}(DC\_CAN_k) + \gamma_{003}(DC\_DOM_k) + \gamma_{004}(DC\_GRC_k) \\ & + \gamma_{005}(DC\_IDN_k) + \gamma_{006}(DC\_IRL_k) + \gamma_{007}(DC\_KOR_k) + \gamma_{008}(DC\_MAR_k) \\ & + \gamma_{009}(DC\_MKD_k) + \gamma_{0010}(DC\_NZL_k) + \gamma_{0011}(DC\_PHL_k) + \gamma_{0012}(DC\_TUR_k) + u_{00k}\end{aligned}$$

Table G11 shows the fixed effects for Model 02. The intercept (i.e., 425) represents the reference point from which to evaluate the now twelve coefficients (e.g., Korea is associated with +103 PISA reading score).

Table G11

Model 02 Fixed Effects

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
For INTRCPT2, β_{00}					
INTRCPT3, γ_{000}	424.529538	4.647458	91.347	29	<0.001
DC_AUS, γ_{001}	69.316977	22.869961	3.031	29	0.005
DC_CAN, γ_{002}	74.762934	22.787569	3.281	29	0.003
DC_DOM, γ_{003}	-142.282564	33.071557	-4.302	29	<0.001
DC_GRC, γ_{004}	-52.532902	25.269703	-2.079	29	0.047
DC_IDN, γ_{005}	-129.277461	30.553139	-4.231	29	<0.001
DC_IRL, γ_{006}	98.585864	24.190391	4.075	29	<0.001
DC_KOR, γ_{007}	103.300827	30.026244	3.440	29	0.002
DC_MAR, γ_{008}	-88.446663	26.252032	-3.369	29	0.002
DC_MKD, γ_{009}	-77.313472	33.177542	-2.330	29	0.027
DC_NZL, γ_{0010}	88.927079	23.454136	3.792	29	<0.001
DC_PHL, γ_{0011}	-124.360332	28.751305	-4.325	29	<0.001
DC_TUR, γ_{0012}	70.585712	30.886323	2.285	29	0.030

Table G12 shows the variance decomposition for Model 02. After adding the 6 additional fixed effects, the destination country variance decreased to 483.

Table G12

Model 02 Variance Components

Final estimation of level-1 and level-2 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1,r_0	54.82793	3006.10242	3472	7217.91290	<0.001
level-1, e	87.05380	7578.36408			

Final estimation of level-3 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/INTRCPT2,u_{00}	21.97516	482.90755	29	216.07663	<0.001

After adding the destination country fixed effects to the previous model the country-level variance value decreased from 10% to 4%, reducing the destination country variance below 5% of the overall variance (i.e., non-significant). Table G13 shows the model variance decomposition.

Table G13

Model Variance Decomposition

Model	Level	Variance	Variance By Level
Model 00	Student	75780	56%
	School	3004	22%
	Country	2971	22%
Model 01	Student	7575	64%
	School	3015	26%
	Country	1206	10%
Model 02	Student	7578	68%
	School	3006	27%
	Country	483	4%

Summary of Stage 1: Twelve Destination Countries Were Identified for Fixed Effects

Three total 3-level hierarchical models (i.e., students nested within schools nested within destination countries) were specified and run for Stage 1: Models 00, 01, and 02 were specified toward identifying destination countries that have an outsized impact on student achievement. In addition, the residuals were examined to identify which destination countries might differ from the rest. The reason for examining residuals was to identify destination countries that may be significantly different from the rest. The goal was to reduce the variance for the destination countries until it becomes small (i.e., non-significant) at less than 5% of the overall variance. Once that goal had been achieved, twelve countries had been designated as fixed effects dummy variables (i.e., Korea, Ireland, New Zealand, Philippines, Dominican Republic, Indonesia, Canada, Turkey, Australia, Greece, North Macedonia, and Morocco). Ultimately, the aim of this Stage 1 process was to be able to control for these outlier countries by transforming them into fixed effects (i.e., predictor variables) at the school-level for the models in Stage 2.

In Stage 1, three models were specified and run towards identifying destination countries that have an outsized impact on student achievement: Model 00, Model 01, and Model 02. First, a null model (i.e., Model 00) was specified and the variance decomposition was examined. In addition, the residuals were examined to identify which destination countries might differ from the rest. These countries would be added into subsequent Stage 1 models (e.g., Model 01, Model 02) as fixed effects dummy variables. The goal was to reduce the variance for the destination countries until it

becomes small (i.e., non-significant) at less than 5% of the overall variance. Once that goal had been achieved, twelve countries had been designated as fixed effects dummy variables (i.e., Korea, Ireland, New Zealand, Philippines, Dominican Republic, Indonesia, Canada, Turkey, Australia, Greece, North Macedonia, and Morocco). The country-level variance value decreased on each Stage 1 model from 22% to 4%, meeting the goal of reducing the variance below 5%.

Ultimately, the aim of this Stage 1 process was to be able to control for these outlier countries by transforming them into fixed effects (i.e., predictor variables) at the student-level for the models in Stage 2. The rationale for adding these countries as level-2 variables is that each school can be reasoned as having characteristics imparted upon it by the country the school is located within. This process solves the issue with complex cross-nesting as it essentially removes destination country as its own level, and instead puts the impact of destination country as a covariate within the model, thereby allowing origin country influences to become the focal point of the modeling.

Stage 2: Modeling Cross-Classified Models

The second stage of the two-stage approach to modeling was to specify a 2-level, cross-classified multilevel model. Students are cross-nested within two different level-2 groups: schools and origin countries. Figure G7 is a visualization of the data as it is to be modeled. The model is level-1 (students) cross-nested within level-2a (schools) and level-2b (origin countries). Note that the destination country is not a level itself but rather fixed effect variables at student-level.

Figure G7

Structure of the Data

Nine total cross-nested models were specified and run for Stage 2: Models 0, 1, and 2 were specified towards creating a baseline model from which to add the predictors of interest in this study. This started with a null model (i.e., Model 0) followed

by two subsequent models, building towards a baseline model. The first subsequent model added the fixed effects countries identified in Stage 1 (i.e., Model 02). The second subsequent model added typical control variables (e.g., sex, socio-economic status). This process formed the baseline model (i.e., Model 3) from which to test the variables of interest against, one at a time.

Then Models 3, 4, 5, and 6 swapped in the main variables of interest for this study (i.e., LD, HDI, GAIN, and FD Ratio), one at a time. LD was added at the student-level while HDI, GAIN, and FD Ratio were at the country-level. Next, Models 7, 8, and 9 tested immigration year versions of the country-level variables (i.e., HDI, GAIN, and FD Ratio). For example, the Immigration Year HDI was a measure of each student's HDI value around their unique time of immigration versus the shared country-wide value for all students from that country. These variables were individually swapped in one at and tested, just as the country-level variables were.

In the end, this modeling procedure identified just one covariate of interest that was statistically significant. This covariate was language distance from Model 3. The other variables were not statistically significant when tested individually.

Model 0: Null Model

First, null Model 0 was specified and run. Figure G8 shows the model specification.

Figure G8

Model 0 Specification

Level-1 Model

$$PV1READ_{ijk} = \pi_{0jk} + e_{ijk}$$

Level-2 Model

$$\pi_{0jk} = \theta_0 + b_{00j} + c_{00k}$$

Mixed Model

$$PV1READ_{ijk} = \theta_0 + b_{00j} + c_{00k} + e_{ijk}$$

Table G15 shows the fixed effects for the null model. In the fixed effects, there are no covariates in the null model so the only relevant term is the intercept (i.e., 421) which represents the grand mean reading score for all students in the data.

Table G15

Model 0 Fixed Effects

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTERCEPT, θ_0	421.413573	7.453922	56.536	4595	<0.001

178

Table G16 shows the variance decomposition for the null model. Most of the variance decomposition belongs to the student-level (i.e. 3625), with the rest at the school-level (i.e., 6509) and origin country-level (i.e., 2518).

Table G16

Model 0 Variance Components

Final estimation of row and level-1 variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTROW,b_{00j}	80.68075	6509.38410	3204	17540.24580	<0.001
level-1, e	60.20657	3624.83149			

Final estimation of column level variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTCOL,c_{00k}	50.17736	2517.76706	73	1913.09861	<0.001

Table G17 shows the model deviance and variance decomposition. The first column indicates the model name. The second column indicates the deviance value. The third column indicates model levels. The fourth column indicates the variance by level. The fourth column indicates the variance decomposition by level. We can interpret the distribution of variance as follows. 29% of the variance is at the individual student-level. 51% of the variance is at the school-level (level-2a). 20% of the variance is at the origin country-level (level-2b). Model 0 is a 2-level cross-nested model. This Model 0 serves as an initial model from which to compare subsequent 2-level cross-nested models against.

Table G17

Model 0 Deviance & Variance Estimates

Model	Deviance	Level	Variance	Variance By Level
Model 0 (Null)	92931	Student	3625	29%
		School	6510	51%
		Country	2518	20%

Model 1: Null Model + Outlier Destination Countries Effects

Next Model 1 was specified and run. Model 1 adds the 12 fixed effects (i.e.,

predictor variables) from Stage 1. Figure G9 shows the model specification.

Figure G9

Model 1 Specification

Level-1 Model

$$PV1READ_{ijk} = \pi_{0jk} + e_{ijk}$$

Level-2 Model

$$\begin{aligned}
\pi_{0jk} = {}& \theta_0 + b_{00j} + c_{00k} \\
& + (\gamma_{01}){*}DC\_AUS_j \\
& + (\gamma_{02}){*}DC\_CAN_j \\
& + (\gamma_{03}){*}DC\_DOM_j \\
& + (\gamma_{04}){*}DC\_GRC_j \\
& + (\gamma_{05}){*}DC\_IDN_j \\
& + (\gamma_{06}){*}DC\_IRL_j \\
& + (\gamma_{07}){*}DC\_KOR_j \\
& + (\gamma_{08}){*}DC\_MAR_j \\
& + (\gamma_{09}){*}DC\_MKD_j \\
& + (\gamma_{010}){*}DC\_NZL_j \\
& + (\gamma_{011}){*}DC\_PHL_j \\
& + (\gamma_{012}){*}DC\_TUR_j
\end{aligned}$$

Table G19 shows the fixed effects for Model 1. The intercept (i.e., 419) represents the reference point from which to evaluate the coefficients of the twelve fixed effects (i.e., Korea, Ireland, New Zealand, Philippines, Dominican Republic, and Indonesia, Canada, Turkey, Australia, Greece, North Macedonia, and Morocco). This means that immigrant students who settled in Korea were associated with scores 141 points higher than the mean score of students attending school in any other country (i.e., 419); except for those twelve countries modeled as fixed effects.

Table G19

Model 1 Fixed Effects

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTERCEPT, θ_0	419.408288	7.503387	55.896	4595	<0.001
DC_AUS, γ_{01}	83.188904	15.272295	5.447	3203	<0.001
DC_CAN, γ_{02}	110.143538	15.752035	6.992	3203	<0.001
DC_DOM, γ_{03}	-119.701021	36.493948	-3.280	3203	0.001
DC_GRC, γ_{04}	22.041708	28.915675	0.762	3203	0.446
DC_IDN, γ_{05}	-146.169197	50.012320	-2.923	3203	0.003
DC_IRL, γ_{06}	88.971781	17.861796	4.981	3203	<0.001
DC_KOR, γ_{07}	141.339045	34.430355	4.105	3203	<0.001
DC_MAR, γ_{08}	-49.684992	23.007797	-2.159	3203	0.031
DC_MKD, γ_{09}	-37.563209	28.857803	-1.302	3203	0.193
DC_NZL, γ_{010}	108.389342	18.010312	6.018	3203	<0.001
DC_PHL, γ_{011}	-85.933752	33.982547	-2.529	3203	0.011
DC_TUR, γ_{012}	1.632383	32.278151	0.051	3203	0.960

Table G20 shows the variance decomposition for Model 1. After adding the twelve fixed effects, the origin country-level variance decreased to 1585.

Table G20

Model 1 Variance Components

Final estimation of row and level-1 variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTROW,b_{00j}	72.14237	5204.52117	3204	14235.74642	<0.001
level-1, e	60.31211	3637.55066			

Final estimation of column level variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTCOL,c_{00k}	39.80890	1584.74875	73	651.19026	<0.001

Table G21 compares the deviance and variance estimates between models, which is used to assess the change in goodness of fit between models. M1 deviance lowered by approximately 575 from 92931 to 92355. A lower deviance suggests a better model fit. A likelihood ratio test confirmed the change in deviance was statistically significant (p=<0.001). Additionally, the variance decreased at all levels.

Table G21

Model 1 Deviance & Variance Estimates Compared to Model 0

Model	Deviance	Level	Variance	Variance By Level
Model 0 (Null)	92931	Student	3625	29%
		School	6510	51%
		Country	2518	20%
Model 1 (Null + Stage 1 Effects)	92355*	Student	3638	35%
		School	5205	50%
		Country	1585	15%
*Note. A*sterisk after the model name denotes the model deviance is statistically significantly different from the baseline Model 0. **Bold text** denotes models with covariates that were statistically significant within the model.				

Model 2: Null Model + Controls = (Baseline Model)

Next Model 2 was specified and run. Model 2 adds control variables for student

socio-economic status and student sex, as well as for school-level socio-economic

status. Figure G10 shows the model specification.

Figure G10

Model 2 Specification

Level-1 Model

$$PV1READ_{ijk} = \pi_{0jk} + \pi_{1jk}*(SEX2_{ijk}) + \pi_{2jk}*(ESCS_{ijk}) + e_{ijk}$$

Level-2 Model

$$\pi_{0jk} = \theta_0 + b_{00j} + c_{00k}$$
$$+ (\gamma_{01})*SCHSES_j$$
$$+ (\gamma_{02})*DC\_AUS_j$$
$$+ (\gamma_{03})*DC\_CAN_j$$
$$+ (\gamma_{04})*DC\_DOM_j$$
$$+ (\gamma_{05})*DC\_GRC_j$$
$$+ (\gamma_{06})*DC\_IDN_j$$
$$+ (\gamma_{07})*DC\_IRL_j$$
$$+ (\gamma_{08})*DC\_KOR_j$$
$$+ (\gamma_{09})*DC\_MAR_j$$
$$+ (\gamma_{010})*DC\_MKD_j$$
$$+ (\gamma_{011})*DC\_NZL_j$$
$$+ (\gamma_{012})*DC\_PHL_j$$
$$+ (\gamma_{013})*DC\_TUR_j$$
$$\pi_{1jk} = \theta_1$$
$$\pi_{2jk} = \theta_2$$

185

Table G23 shows the fixed effects for Model 2. The intercept (i.e., 468) represents the reference point from which to evaluate the coefficients.

Table G23

Model 2 Fixed Effects

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTERCEPT,θ_0	468.198606	9.416873	49.719	4595	<0.001
SCHSES, γ_{01}	-0.712473	0.218751	-3.257	3203	0.001
DC_AUS, γ_{02}	62.912749	13.463889	4.673	3203	<0.001
DC_CAN, γ_{03}	83.103380	13.701411	6.065	3203	<0.001
DC_DOM, γ_{04}	-124.407954	33.721115	-3.689	3203	<0.001
DC_GRC, γ_{05}	11.472375	26.905445	0.426	3203	0.670
DC_IDN, γ_{06}	-133.803644	45.969931	-2.911	3203	0.004
DC_IRL, γ_{07}	79.353463	15.776611	5.030	3203	<0.001
DC_KOR, γ_{08}	102.857786	31.658995	3.249	3203	0.001
DC_MAR, γ_{09}	-29.137537	22.198247	-1.313	3203	0.189
DC_MKD, γ_{010}	-52.981074	27.001246	-1.962	3203	0.050
DC_NZL, γ_{011}	85.539679	16.282389	5.254	3203	<0.001
DC_PHL, γ_{012}	-86.610729	31.136779	-2.782	3203	0.005
DC_TUR, γ_{013}	8.247603	30.279554	0.272	3203	0.785
For SEX2, π_1					
INTERCEPT,θ_1	-37.623903	8.841099	-4.256	4595	<0.001
For ESCS, π_2					
INTERCEPT,θ_2	12.996063	4.661383	2.788	4595	0.005

Table G24 shows the variance decomposition for Model 2. After adding the additional covariates, the origin country-level variance decreased to 1112.

Table G24

Model 2 Variance Components

Final estimation of row and level-1 variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTROW,b_{00j}	65.52090	4292.98864	3191	12676.67736	<0.001
level-1, e	59.49339	3539.46358			

Final estimation of column level variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTCOL,c_{00k}	33.34150	1111.65560	60	739.02206	<0.001

Table G25 compares the deviance and variance estimates between models, which is used to assess the change in goodness of fit between models. M2 deviance lowered by approximately 635 from 92355 to 91720. A lower deviance suggests a better model fit. A likelihood ratio test confirmed the change in deviance was statistically significant (p=<0.001). Additionally, the variance decreased at all levels.

Table G25

Model 2 Deviance & Variance Estimates Compared to Model 1

Model	Deviance	Level	Variance	Variance By Level
Model 1 (Null + Stage 1 Effects)	92355	Student	3638	35%
		School	5205	50%
		Country	1585	15%
Model 2 (Null + Stage 1 Effects + Controls)	91720*	Student	3539	40%
		School	4293	48%
		Country	1112	12%
*Note. A*sterisk after the model name denotes the model deviance is statistically significantly different from the Model 1. **Bold text** denotes models with covariates that were statistically significant within the model.				

Since Model 2 includes both the fixed effects from Stage 1 and the control variables, this model serves as the baseline model from which to test the various coefficients of interest to the study.

Model 3: Language Distance

Next Model 3 was specified and run. Model 3 adds the first student-level

predictor of interest for this study: language distance. Figure G11 shows the model

specification.

Figure G11

Model 3 Specification

Level-1 Model

$$PV1READ_{ijk} = \pi_{0jk} + \pi_{1jk}*(SEX2_{ijk}) + \pi_{2jk}*(ESCS_{ijk}) + \pi_{3jk}*(LANGDIST_{ijk}) + e_{ijk}$$

Level-2 Model

$$\pi_{0jk} = \theta_0 + b_{00j} + c_{00k}$$
$$+ (\gamma_{01})*SCHSES_j$$
$$+ (\gamma_{02})*DC\_AUS_j$$
$$+ (\gamma_{03})*DC\_CAN_j$$
$$+ (\gamma_{04})*DC\_DOM_j$$
$$+ (\gamma_{05})*DC\_GRC_j$$
$$+ (\gamma_{06})*DC\_IDN_j$$
$$+ (\gamma_{07})*DC\_IRL_j$$
$$+ (\gamma_{08})*DC\_KOR_j$$
$$+ (\gamma_{09})*DC\_MAR_j$$
$$+ (\gamma_{010})*DC\_MKD_j$$
$$+ (\gamma_{011})*DC\_NZL_j$$
$$+ (\gamma_{012})*DC\_PHL_j$$
$$+ (\gamma_{013})*DC\_TUR_j$$
$$\pi_{1jk} = \theta_1$$
$$\pi_{2jk} = \theta_2$$
$$\pi_{3jk} = \theta_3$$

Table G27 shows the fixed effects for Model 3. The intercept (i.e., 476) represents the reference point from which to evaluate the coefficients. The student-level language distance measure was statistically significant (-0.31, SE=0.13, p=0.022). This means that a 1 unit change in language distance is associated with -0.31 change in PISA reading score).

Table G27

Model 3 Fixed Effects

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTERCEPT, θ_0	475.706107	10.315178	46.117	4595	<0.001
SCHSES, γ_{01}	-0.679852	0.217885	-3.120	3203	0.002
DC_AUS, γ_{02}	64.392495	13.339437	4.827	3203	<0.001
DC_CAN, γ_{03}	79.418789	13.796068	5.757	3203	<0.001
DC_DOM, γ_{04}	-122.781979	33.519623	-3.663	3203	<0.001
DC_GRC, γ_{05}	9.350883	26.667683	0.351	3203	0.726
DC_IDN, γ_{06}	-134.962667	45.652022	-2.956	3203	0.003
DC_IRL, γ_{07}	79.349199	15.663576	5.066	3203	<0.001
DC_KOR, γ_{08}	105.257943	31.457890	3.346	3203	<0.001
DC_MAR, γ_{09}	-40.201020	22.462971	-1.790	3203	0.074
DC_MKD, γ_{010}	-60.674767	26.982879	-2.249	3203	0.025
DC_NZL, γ_{011}	85.285164	16.159834	5.278	3203	<0.001
DC_PHL, γ_{012}	-83.232606	31.024108	-2.683	3203	0.007
DC_TUR, γ_{013}	0.993082	30.220247	0.033	3203	0.974
For SEX2, π_1					
INTERCEPT, θ_1	-35.802803	8.637343	-4.145	4595	<0.001
For ESCS, π_2					
INTERCEPT, θ_2	13.114632	4.630827	2.832	4595	0.005
For LANGDIST, π_3					
INTERCEPT, θ_3	-0.307951	0.134811	-2.284	4595	0.022

Table G28 shows the variance decomposition for each level of Model 3.

Table G28

Model 3 Variance Components

Final estimation of row and level-1 variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTROW,b_{00j}	65.32497	4267.35168	3191	12762.09815	<0.001
level-1, e	59.06257	3488.38747			

Final estimation of column level variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTCOL,c_{00k}	32.84085	1078.52127	60	664.82598	<0.001

Table G29 compares the deviance and variance estimates between models, which is used to assess the change in goodness of fit between models. Model 3 deviance lowered by approximately 95 from 91720 to 91625. A lower deviance suggests a better model fit. A likelihood ratio test confirmed the change in deviance was statistically significant (p=<0.001). Additionally, the variance decreased at all levels.

Table G29

Model 3 Deviance & Variance Estimates Compared to Baseline

Model	Deviance	Level	Variance	Variance By Level
Model 2 (Null + Stage 1 Effects + Controls)	91720	Student	3539	40%
		School	4293	48%
		Country	1112	12%
Model 3 (Language Distance)	**91625***	**Student**	**3488**	**39%**
		School	**4267**	**48%**
		Country	**1079**	**12%**
*Note. A*sterisk after the model name denotes the model deviance is statistically significantly different from the baseline Model 2. **Bold text** denotes models with covariates that were statistically significant within the model.				

Model 4: HDI

Next Model 4 was specified and run. Model 4 adds the first origin country predictor of interest for this study: Human Development Index (HDI). Figure G12 shows the model specification.

Figure G12

Model 4 Specification

Level-1 Model

$$PV1READ_{ijk} = \pi_{0jk} + \pi_{1jk}*(SEX2_{ijk}) + \pi_{2jk}*(ESCS_{ijk}) + e_{ijk}$$

Level-2 Model

$$
\begin{aligned}
\pi_{0jk} = {}& \theta_0 + b_{00j} + c_{00k} \\
& + (\gamma_{01})*SCHSES_j \\
& + (\gamma_{02})*DC\_AUS_j \\
& + (\gamma_{03})*DC\_CAN_j \\
& + (\gamma_{04})*DC\_DOM_j \\
& + (\gamma_{05})*DC\_GRC_j \\
& + (\gamma_{06})*DC\_IDN_j \\
& + (\gamma_{07})*DC\_IRL_j \\
& + (\gamma_{08})*DC\_KOR_j \\
& + (\gamma_{09})*DC\_MAR_j \\
& + (\gamma_{010})*DC\_MKD_j \\
& + (\gamma_{011})*DC\_NZL_j \\
& + (\gamma_{012})*DC\_PHL_j \\
& + (\gamma_{013})*DC\_TUR_j \\
& + (\beta_{01})*HDI_k \\
\pi_{1jk} = {}& \theta_1 \\
\pi_{2jk} = {}& \theta_2
\end{aligned}
$$

Table G31 shows the fixed effects for Model 4. The intercept (i.e., 425) represents the reference value from which to evaluate the coefficients. The country-level HDI was not statistically significant (57, SE=41, p=0.165).

Table G31

Model 4 Fixed Effects

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTERCEPT,θ_0	424.769955	31.751498	13.378	4595	<0.001
SCHSES, γ_{01}	-0.711296	0.218636	-3.253	3203	0.001
DC_AUS, γ_{02}	62.741993	13.427020	4.673	3203	<0.001
DC_CAN, γ_{03}	83.302326	13.674313	6.092	3203	<0.001
DC_DOM, γ_{04}	-122.564218	33.728158	-3.634	3203	<0.001
DC_GRC, γ_{05}	11.970486	26.824529	0.446	3203	0.655
DC_IDN, γ_{06}	-135.109273	45.924018	-2.942	3203	0.003
DC_IRL, γ_{07}	78.887769	15.752828	5.008	3203	<0.001
DC_KOR, γ_{08}	102.458974	31.647751	3.237	3203	0.001
DC_MAR, γ_{09}	-29.778180	22.218522	-1.340	3203	0.180
DC_MKD, γ_{010}	-52.528328	26.986426	-1.946	3203	0.052
DC_NZL, γ_{011}	85.209228	16.245958	5.245	3203	<0.001
DC_PHL, γ_{012}	-87.220627	31.133350	-2.802	3203	0.005
DC_TUR, γ_{013}	7.689867	30.280406	0.254	3203	0.800
HDI, β_{014}	57.517213	41.013690	1.402	72	0.165
For SEX2, π_1					
INTERCEPT,θ_1	-37.614290	8.840099	-4.255	4595	<0.001
For ESCS, π_2					
INTERCEPT,θ_2	12.915532	4.676950	2.762	4595	0.006

Table G32 shows the variance decomposition for each level of Model 4.

Table G32

Model 4 Variance Components

Final estimation of row and level-1 variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTROW,b_{00j}	65.50263	4290.59429	3190	12661.38167	<0.001
level-1, e	59.49329	3539.45178			

Final estimation of column level variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTCOL,c_{00k}	32.99243	1088.50068	59	739.52749	<0.001

Table G33 compares the deviance and variance estimates between models, which is used to assess the change in goodness of fit between models. Model 4 deviance lowered by approximately 2 from 91720 to 91718. A lower deviance suggests a better model fit. A likelihood ratio test confirmed the change in deviance was not statistically significant (p=0.113). Additionally, the variance decreased at all levels.

Table G33

Model 4 Deviance & Variance Estimates Compared to Baseline

Model	Deviance	Level	Variance	Variance By Level
Model 2 (Null + Stage 1 Effects + Controls)	91720	Student	3539	40%
		School	4293	48%
		Country	1112	12%
Model 4 (HDI)	91718	Student	3539	40%
		School	4291	48%
		Country	1089	12%
*Note. A*sterisk after the model name denotes the model deviance is statistically significantly different from the baseline Model 2. **Bold text** denotes models with covariates that were statistically significant within the model.				

Model 5: GAIN

Next Model 5 was specified and run. Model 5 swaps in the second origin country predictor of interest (i.e., Notre Dame-Global Adaptation Initiative (GAIN)) while swapping out the first (HDI). Figure G13 shows the model specification.

Figure G13

Model 5 Specification

Level-1 Model

$$PV1READ_{ijk} = \pi_{0jk} + \pi_{1jk}*(SEX2_{ijk}) + \pi_{2jk}*(ESCS_{ijk}) + e_{ijk}$$

Level-2 Model

$$\pi_{0jk} = \theta_0 + b_{00j} + c_{00k}$$
$$+ (\gamma_{01})*SCHSES_j$$
$$+ (\gamma_{02})*DC\_AUS_j$$
$$+ (\gamma_{03})*DC\_CAN_j$$
$$+ (\gamma_{04})*DC\_DOM_j$$
$$+ (\gamma_{05})*DC\_GRC_j$$
$$+ (\gamma_{06})*DC\_IDN_j$$
$$+ (\gamma_{07})*DC\_IRL_j$$
$$+ (\gamma_{08})*DC\_KOR_j$$
$$+ (\gamma_{09})*DC\_MAR_j$$
$$+ (\gamma_{010})*DC\_MKD_j$$
$$+ (\gamma_{011})*DC\_NZL_j$$
$$+ (\gamma_{012})*DC\_PHL_j$$
$$+ (\gamma_{013})*DC\_TUR_j$$
$$+ (\beta_{01})*GAIN_k$$
$$\pi_{1jk} = \theta_1$$
$$\pi_{2jk} = \theta_2$$

197

Table G35 shows the fixed effects for Model 5. The intercept (i.e., 420) represents the reference value from which to evaluate the coefficients. The country-level GAIN coefficient was not statistically significant (0.92, SE=0.52, p=0.085).

Table G35

Model 5 Deviance & Variance Estimates Compared to Baseline

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTERCEPT, θ_0	419.896367	28.523272	14.721	4595	<0.001
SCHSES, γ_{01}	-0.711069	0.218591	-3.253	3203	0.001
DC_AUS, γ_{02}	61.812720	13.458229	4.593	3203	<0.001
DC_CAN, γ_{03}	82.880352	13.663116	6.066	3203	<0.001
DC_DOM, γ_{04}	-122.146740	33.688531	-3.626	3203	<0.001
DC_GRC, γ_{05}	13.145339	26.802279	0.490	3203	0.624
DC_IDN, γ_{06}	-135.804801	45.908295	-2.958	3203	0.003
DC_IRL, γ_{07}	77.834479	15.815783	4.921	3203	<0.001
DC_KOR, γ_{08}	102.117437	31.655056	3.226	3203	0.001
DC_MAR, γ_{09}	-30.031615	22.220050	-1.352	3203	0.177
DC_MKD, γ_{010}	-51.652248	27.001423	-1.913	3203	0.056
DC_NZL, γ_{011}	84.336684	16.289651	5.177	3203	<0.001
DC_PHL, γ_{012}	-87.312620	31.098080	-2.808	3203	0.005
DC_TUR, γ_{013}	7.369153	30.278995	0.243	3203	0.808
GAIN, β_{014}	0.916110	0.524856	1.745	72	0.085
For SEX2, π_1					
INTERCEPT, θ_1	-37.605878	8.839883	-4.254	4595	<0.001
For ESCS, π_2					
INTERCEPT, θ_2	12.910074	4.675877	2.761	4595	0.006

Table G36 shows the variance decomposition for each level of Model 5.

Table G36

Model 5 Variance Components

Final estimation of row and level-1 variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTROW,b_{00j}	65.49239	4289.25327	3190	12660.55995	<0.001
level-1, e	59.49375	3539.50608			

Final estimation of column level variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTCOL,c_{00k}	32.78801	1075.05391	59	748.80398	<0.001

Table G37 compares the deviance and variance estimates between models, which is used to assess the change in goodness of fit between models. Model 5 deviance lowered by approximately 4 from 91720 to 91716. A lower deviance suggests a better model fit. A likelihood ratio test confirmed the change in deviance was statistically significant (p=<0.049). Additionally, the variance decreased at school- and country-levels yet remained practically the same at student-level.

Table G37

Model 5 Deviance & Variance Estimates Compared to Baseline

Model	Deviance	Level	Variance	Variance By Level
Model 2 (Null + Stage 1 Effects + Controls)	91720	Student	3539	40%
		School	4293	48%
		Country	1112	12%
Model 5 (GAIN)	91716*	Student	3540	40%
		School	4289	48%
		Country	1075	12%
Note. Asterisk after the model name denotes the model deviance is statistically significantly different from the baseline Model 2. **Bold text** denotes models with covariates that were statistically significant within the model.				

Model 6: FD Ratio

Next Model 6 was specified and run. Model 6 swaps in the final origin country predictor of interest (i.e., FD Ratio while swapping out the previous (GAIN). Figure G14 shows the model specification.

Figure G14

Model 6 Specification

Level-1 Model

$$PV1READ_{ijk} = \pi_{0jk} + \pi_{1jk}*(SEX2_{ijk}) + \pi_{2jk}*(ESCS_{ijk}) + e_{ijk}$$

Level-2 Model

$$
\begin{aligned}
\pi_{0jk} = {}& \theta_0 + b_{00j} + c_{00k} \\
& + (\gamma_{01})*SCHSES_j \\
& + (\gamma_{02})*DC\_AUS_j \\
& + (\gamma_{03})*DC\_CAN_j \\
& + (\gamma_{04})*DC\_DOM_j \\
& + (\gamma_{05})*DC\_GRC_j \\
& + (\gamma_{06})*DC\_IDN_j \\
& + (\gamma_{07})*DC\_IRL_j \\
& + (\gamma_{08})*DC\_KOR_j \\
& + (\gamma_{09})*DC\_MAR_j \\
& + (\gamma_{010})*DC\_MKD_j \\
& + (\gamma_{011})*DC\_NZL_j \\
& + (\gamma_{012})*DC\_PHL_j \\
& + (\gamma_{013})*DC\_TUR_j \\
& + (\beta_{01})*DISPLACE_k \\
\pi_{1jk} = {}& \theta_1 \\
\pi_{2jk} = {}& \theta_2
\end{aligned}
$$

Table G39 shows the fixed effects for Model 6. The intercept (i.e., 468) represents the reference value from which to evaluate the coefficients. The country-level FD Ratio coefficient was not statistically significant (9.26, SE=4.89, p=0.062).

Table G39

Model 6 Fixed Effects

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTERCEPT,θ_0	468.365840	9.349413	50.096	4595	<0.001
SCHSES, γ_{01}	-0.709588	0.218644	-3.245	3203	0.001
DC_AUS, γ_{02}	62.742145	13.345615	4.701	3203	<0.001
DC_CAN, γ_{03}	83.235510	13.625330	6.109	3203	<0.001
DC_DOM, γ_{04}	-123.888397	33.610272	-3.686	3203	<0.001
DC_GRC, γ_{05}	11.426718	26.652769	0.429	3203	0.668
DC_IDN, γ_{06}	-135.288266	45.594632	-2.967	3203	0.003
DC_IRL, γ_{07}	78.600935	15.669153	5.016	3203	<0.001
DC_KOR, γ_{08}	102.960078	31.597882	3.258	3203	0.001
DC_MAR, γ_{09}	-29.708331	22.189826	-1.339	3203	0.181
DC_MKD, γ_{010}	-52.563822	26.944110	-1.951	3203	0.051
DC_NZL, γ_{011}	84.917421	16.151248	5.258	3203	<0.001
DC_PHL, γ_{012}	-86.413145	31.031531	-2.785	3203	0.005
DC_TUR, γ_{013}	7.385495	30.256491	0.244	3203	0.807
DISPLACE, β_{014}	9.263301	4.893082	1.893	72	0.062
For SEX2, π_1					
INTERCEPT,θ_1	-37.543937	8.839519	-4.247	4595	<0.001
For ESCS, π_2					
INTERCEPT,θ_2	12.927188	4.667828	2.769	4595	0.006

Table G40 shows the variance decomposition for each level of Model 6.

Table G40

Model 6 Variance Components

Final estimation of row and level-1 variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTROW,b_{00j}	65.54439	4296.06695	3190	12671.76669	<0.001
level-1, e	59.48145	3538.04334			

Final estimation of column level variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTCOL,c_{00k}	32.26907	1041.29313	59	723.73446	<0.001

Table G41 compares the deviance and variance estimates between models, which is used to assess the change in goodness of fit between models. Model 6 deviance lowered by approximately 4 from 91720 to 91716. A lower deviance suggests a better model fit. A likelihood ratio test confirmed the change in deviance was statistically significant (p=<0.046). Additionally, the variance decreased at all the school- and country-levels.

Table G41

Model 6 Deviance & Variance Estimates Compared to Baseline

Model	Deviance	Level	Variance	Variance By Level
Model 2 (Null + Stage 1 Effects + Controls)	91720	Student	3539	40%
		School	4293	48%
		Country	1112	12%
Model 6 (FD Ratio)	91716*	Student	3538	40%
		School	4296	48%
		Country	1041	12%
Note. Asterisk after the model name denotes the model deviance is statistically significantly different from the baseline Model 2. **Bold text** denotes models with covariates that were statistically significant within the model.				

Model 7: HDI (Immigration Year)

Next Model 7 was specified and run. Model 7 tests immigration year versions of the country-level variables. First was Immigration Year HDI. Figure G15 shows the model specification.

Figure G15

Model 7 Specification

Level-1 Model

$$PV1READ_{ijk} = \pi_{0jk} + \pi_{1jk}*(SEX2_{ijk}) + \pi_{2jk}*(ESCS_{ijk}) + \pi_{3jk}*(IY\_HDI_{ijk}) + e_{ijk}$$

Level-2 Model

$$
\begin{aligned}
\pi_{0jk} = {} & \theta_0 + b_{00j} + c_{00k} \\
& + (\gamma_{01})*SCHSES_j \\
& + (\gamma_{02})*DC\_AUS_j \\
& + (\gamma_{03})*DC\_CAN_j \\
& + (\gamma_{04})*DC\_DOM_j \\
& + (\gamma_{05})*DC\_GRC_j \\
& + (\gamma_{06})*DC\_IDN_j \\
& + (\gamma_{07})*DC\_IRL_j \\
& + (\gamma_{08})*DC\_KOR_j \\
& + (\gamma_{09})*DC\_MAR_j \\
& + (\gamma_{010})*DC\_MKD_j \\
& + (\gamma_{011})*DC\_NZL_j \\
& + (\gamma_{012})*DC\_PHL_j \\
& + (\gamma_{013})*DC\_TUR_j \\
\pi_{1jk} = {} & \theta_1 \\
\pi_{2jk} = {} & \theta_2 \\
\pi_{3jk} = {} & \theta_3
\end{aligned}
$$

Table G43 shows the fixed effects for Model 7. The intercept (i.e., 680) represents the reference point from which to evaluate the coefficients. While Immigration Year HDI had a statistically significant p-value (coefficient = -283, SE=106, p=0.007), there were two other indicators that suggested not interpreting this covariate due to possible collinearity (i.e., a positive/negative sign switch and an increased standard error). First the HDI coefficient sign changed from positive (+58) in the country-level to negative (-283) in the immigration year version of HDI. Second was that the standard error increased from the country-level value (21) to a value more than 5% larger (106) in the Immigration Year HDI. To further investigate, simpler, non-multilevel regressions were tested with both the country and student HDI variable to see if the unusual patterns were also found in a simpler model. In those cases both HDI variables had coefficients with positive signs and similar standard errors, suggesting that the immigration year, cross-classified HDI was statistically problematic and therefore Model 7 was not interpretable.

Table G43

Model 7 Fixed Effects

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTERCEPT, θ_0	679.551646	82.461993	8.241	4595	<0.001
SCHSES, γ_{01}	-0.674344	0.213896	-3.153	3203	0.002
DC_AUS, γ_{02}	67.826354	14.965559	4.532	3203	<0.001
DC_CAN, γ_{03}	84.560475	14.577480	5.801	3203	<0.001
DC_DOM, γ_{04}	-121.196870	35.518813	-3.412	3203	<0.001
DC_GRC, γ_{05}	12.280466	30.244981	0.406	3203	0.685
DC_IDN, γ_{06}	-132.349200	49.443746	-2.677	3203	0.007
DC_IRL, γ_{07}	81.956432	17.025315	4.814	3203	<0.001
DC_KOR, γ_{08}	110.631209	32.335228	3.421	3203	<0.001
DC_MAR, γ_{09}	-29.489373	22.228397	-1.327	3203	0.185
DC_MKD, γ_{010}	-53.195640	27.844359	-1.910	3203	0.056
DC_NZL, γ_{011}	91.478840	18.054322	5.067	3203	<0.001
DC_PHL, γ_{012}	-76.479222	32.453003	-2.357	3203	0.019
DC_TUR, γ_{013}	4.521266	30.569614	0.148	3203	0.882
For SEX2, π_1					
INTERCEPT, θ_1	-36.580743	8.668222	-4.220	4595	<0.001
For ESCS, π_2					
INTERCEPT, θ_2	12.975846	4.634236	2.800	4595	0.005
For IY_HDI, π_3					
INTERCEPT, θ_3	-283.167444	105.570377	-2.682	4595	0.007

Table G44 shows the variance decomposition for each level of Model 7.

Table G44

Model 7 Variance Components

Final estimation of row and level-1 variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTROW,b_{00j}	64.56706	4168.90540	3191	12675.87913	<0.001
level-1, e	59.28936	3515.22823			

Final estimation of column level variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTCOL,c_{00k}	55.20107	3047.15859	60	1724.20361	<0.001

Table G45 compares the deviance and variance estimates between models, which is used to assess the change in goodness of fit between models. Model 7 deviance lowered by approximately 48 from 91720 to 91672. A lower deviance suggests a better model fit. A likelihood ratio test confirmed the change in deviance was statistically significant (p=<0.001). Additionally, the variance decreased at student- and school-levels while the country-level increased by over 2000.

Table G45

Model 7 Deviance & Variance Estimates Compared to Baseline

Model	Deviance	Level	Variance	Variance By Level
Model 2 (Null + Stage 1 Effects + Controls)	91720	Student	3539	40%
		School	4293	48%
		Country	1112	12%
Model 7 (Immigration Year HDI)	91672*	Student	3515	33%
		School	4169	39%
		Country	3047	28%
Note. Asterisk after the model name denotes the model deviance is statistically significantly different from the baseline Model 2. **Bold text** denotes models with covariates that were statistically significant within the model.				

Model 8: GAIN (Immigration Year)

Next Model 8 was specified and run. Model 8 tests immigration year versions of the country-level variables. This was Immigration Year GAIN. Figure G16 shows the model specification.

Figure G16

Model 8 Specification

Level-1 Model

$$PV1READ_{ijk} = \pi_{0jk} + \pi_{1jk}*(SEX2_{ijk}) + \pi_{2jk}*(ESCS_{ijk}) + \pi_{3jk}*(IY\_GAIN_{ijk}) + e_{ijk}$$

Level-2 Model

$$\begin{aligned}
\pi_{0jk} = \theta_0 &+ b_{00j} + c_{00k} \\
&+ (\gamma_{01})*SCHSES_j \\
&+ (\gamma_{02})*DC\_AUS_j \\
&+ (\gamma_{03})*DC\_CAN_j \\
&+ (\gamma_{04})*DC\_DOM_j \\
&+ (\gamma_{05})*DC\_GRC_j \\
&+ (\gamma_{06})*DC\_IDN_j \\
&+ (\gamma_{07})*DC\_IRL_j \\
&+ (\gamma_{08})*DC\_KOR_j \\
&+ (\gamma_{09})*DC\_MAR_j \\
&+ (\gamma_{010})*DC\_MKD_j \\
&+ (\gamma_{011})*DC\_NZL_j \\
&+ (\gamma_{012})*DC\_PHL_j \\
&+ (\gamma_{013})*DC\_TUR_j
\end{aligned}$$
$$\pi_{1jk} = \theta_1$$
$$\pi_{2jk} = \theta_2$$
$$\pi_{3jk} = \theta_3$$

Table G47 shows the fixed effects for Model 8. The intercept (i.e., 591) represents the reference point from which to evaluate the coefficients. The Immigration Year GAIN coefficient was not statistically significant (-2.36, SE=1.45, p=0.103).

Table G47

Model 8 Fixed Effects

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTERCEPT, θ_0	591.235460	79.442277	7.442	4595	<0.001
SCHSES, γ_{01}	-0.697927	0.216384	-3.225	3203	0.001
DC_AUS, γ_{02}	68.567687	14.834907	4.622	3203	<0.001
DC_CAN, γ_{03}	85.381951	14.439595	5.913	3203	<0.001
DC_DOM, γ_{04}	-124.142485	34.934254	-3.554	3203	<0.001
DC_GRC, γ_{05}	12.771820	29.438205	0.434	3203	0.664
DC_IDN, γ_{06}	-130.254116	48.647117	-2.678	3203	0.007
DC_IRL, γ_{07}	83.988853	16.935996	4.959	3203	<0.001
DC_KOR, γ_{08}	109.868035	32.247090	3.407	3203	<0.001
DC_MAR, γ_{09}	-27.856568	22.321122	-1.248	3203	0.212
DC_MKD, γ_{010}	-52.836706	27.710358	-1.907	3203	0.057
DC_NZL, γ_{011}	91.245406	17.809573	5.123	3203	<0.001
DC_PHL, γ_{012}	-81.046418	32.121186	-2.523	3203	0.012
DC_TUR, γ_{013}	6.472291	30.567710	0.212	3203	0.832
For SEX2, π_1					
INTERCEPT, θ_1	-36.789508	8.600156	-4.278	4595	<0.001
For ESCS, π_2					
INTERCEPT, θ_2	12.928509	4.632487	2.791	4595	0.005
For IY_GAIN, π_3					
INTERCEPT, θ_3	-2.363986	1.449610	-1.631	4595	0.103

Table G48 shows the variance decomposition for each level of Model 8.

Table G48

Model 8 Variance Components

Final estimation of row and level-1 variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTROW,b_{00j}	65.39016	4275.87354	3191	12813.05918	<0.001
level-1, e	59.21287	3506.16409			

Final estimation of column level variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTCOL,c_{00k}	45.59130	2078.56649	60	1287.96924	<0.001

Table G49 compares the deviance and variance estimates between models, which is used to assess the change in goodness of fit between models. Model deviance lowered by approximately 25 from 91720 to 91695. A lower deviance suggests a better model fit. A likelihood ratio test confirmed the change in deviance was statistically significant (p=<0.001). Additionally, the variance decreased at school-level only.

Table G49

Model 8 Deviance & Variance Estimates Compared to Baseline

Model	Deviance	Level	Variance	Variance By Level
Model 2 (Null + Stage 1 Effects + Controls)	91720	Student	3539	40%
		School	4293	48%
		Country	1112	12%
Model 8 (Immigration Year GAIN)	91695*	Student	3506	36%
		School	4276	43%
		Country	2079	21%
*Note. A*sterisk after the model name denotes the model deviance is statistically significantly different from the baseline Model 2. **Bold text** denotes models with covariates that were statistically significant within the model.				

Model 9: FD Ratio (Immigration Year)

Next Model 9 was specified and run. Model 9 tests immigration year versions of the country-level variables. This was the Immigration Year FD Ratio. Figure G17 shows the model specification.

Figure G17

Model 9 Specification

Level-1 Model

$$PV1READ_{ijk} = \pi_{0jk} + \pi_{1jk}*(SEX2_{ijk}) + \pi_{2jk}*(ESCS_{ijk}) + \pi_{3jk}*(IY\_DISPL_{ijk}) + e_{ijk}$$

Level-2 Model

$$\begin{aligned}
\pi_{0jk} =\ & \theta_0 + b_{00j} + c_{00k} \\
& + (\gamma_{01})*SCHSES_j \\
& + (\gamma_{02})*DC\_AUS_j \\
& + (\gamma_{03})*DC\_CAN_j \\
& + (\gamma_{04})*DC\_DOM_j \\
& + (\gamma_{05})*DC\_GRC_j \\
& + (\gamma_{06})*DC\_IDN_j \\
& + (\gamma_{07})*DC\_IRL_j \\
& + (\gamma_{08})*DC\_KOR_j \\
& + (\gamma_{09})*DC\_MAR_j \\
& + (\gamma_{010})*DC\_MKD_j \\
& + (\gamma_{011})*DC\_NZL_j \\
& + (\gamma_{012})*DC\_PHL_j \\
& + (\gamma_{013})*DC\_TUR_j \\
\pi_{1jk} =\ & \theta_1 \\
\pi_{2jk} =\ & \theta_2 \\
\pi_{3jk} =\ & \theta_3
\end{aligned}$$

Table G51 shows the fixed effects for Model 9. The intercept (i.e., 468) represents the reference value from which to evaluate the coefficients. The Immigration Year FD Ratio coefficient was not statistically significant (4.11, SE=4.17, p=0.324).

Table G51

Model 9 Fixed Effects

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTERCEPT, θ_0	467.681564	9.316602	50.199	4595	<0.001
SCHSES, γ_{01}	-0.713120	0.218881	-3.258	3203	0.001
DC_AUS, γ_{02}	63.636645	13.363887	4.762	3203	<0.001
DC_CAN, γ_{03}	83.986596	13.626566	6.163	3203	<0.001
DC_DOM, γ_{04}	-123.404506	33.668752	-3.665	3203	<0.001
DC_GRC, γ_{05}	11.633724	26.725469	0.435	3203	0.663
DC_IDN, γ_{06}	-133.077187	45.769173	-2.908	3203	0.004
DC_IRL, γ_{07}	79.655355	15.695897	5.075	3203	<0.001
DC_KOR, γ_{08}	103.655920	31.602312	3.280	3203	0.001
DC_MAR, γ_{09}	-28.049414	22.203284	-1.263	3203	0.207
DC_MKD, γ_{010}	-52.660097	26.973055	-1.952	3203	0.051
DC_NZL, γ_{011}	86.707005	16.188117	5.356	3203	<0.001
DC_PHL, γ_{012}	-85.537612	31.091260	-2.751	3203	0.006
DC_TUR, γ_{013}	9.622453	30.269108	0.318	3203	0.751
For SEX2, π_1					
INTERCEPT, θ_1	-37.480144	8.806412	-4.256	4595	<0.001
For ESCS, π_2					
INTERCEPT, θ_2	12.906389	4.648847	2.776	4595	0.006
For IY_DISPL, π_3					
INTERCEPT, θ_3	4.111698	4.166138	0.987	4595	0.324

Table G52 shows the variance decomposition for each level of Model 3.

Table G52

Model 9 Variance Components

Final estimation of row and level-1 variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTROW,b_{00j}	65.71781	4318.83039	3191	12746.24509	<0.001
level-1, e	59.40064	3528.43547			

Final estimation of column level variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTCOL,c_{00k}	32.52690	1057.99894	60	712.09305	<0.001

Table G53 compares the deviance and variance estimates between models, which is used to assess the change in goodness of fit between models. Model 9 deviance lowered by approximately 5 from 91720 to 91716. A lower deviance suggests a better model fit. A likelihood ratio test confirmed the change in deviance was statistically significant (p=<0.022). Additionally, the variance decreased at the student-level only.

Table G53

Model 9 Deviance & Variance Estimates Compared to Baseline

Model	Deviance	Level	Variance	Variance By Level
Model 2 (Null + Stage 1 Effects + Controls)	91720	Student	3539	40%
		School	4293	48%
		Country	1112	12%
Model 9 (Immigration Year FD Ratio)	91715*	Student	3528	40%
		School	4319	48%
		Country	1058	12%
*Note. A*sterisk after the model name denotes the model deviance is statistically significantly different from the baseline Model 2. **Bold text** denotes models with covariates that were statistically significant within the model.				

Summary of Stage 2: Two Variables Were Significant; One Model Was Preferred

Nine total cross-nested models were specified and run for Stage 2: Models 0, 1, and 2 were specified towards creating a baseline model from which to add the predictors of interest in this study. This started with a null model (i.e., Model 0) followed by two subsequent models, building towards a baseline model. The first subsequent model added the fixed effects countries identified in Stage 1 (i.e., Model 02). The second subsequent model added typical control variables (e.g., sex, socio-economic status). This process formed the baseline model (i.e., Model 3) from which to test the variables of interest against, one at a time.

Then Models 3, 4, 5, and 6 swapped in the main variables of interest for this study (i.e., LD, HDI, GAIN, and FD Ratio), one at a time. LD was added at the student-level while HDI, GAIN, and FD Ratio were at the country-level. Next, Models 7, 8, and 9 tested immigration year versions of the country-level variables (i.e., HDI, GAIN, and FD Ratio). For example, the Immigration Year HDI was a measure of each student's HDI value around their unique time of immigration versus the shared country-wide value for all students from that country. These variables were individually swapped in one at and tested, just as the country-level variables were.

In the end, this modeling procedure identified just one covariate of interest that was statistically significant. This covariate was language distance from Model 3. The other variables were not statistically significant when tested individually. Table G54 compares all models in Stage 2 at once.

Table G54

Models with Significant Covariates Compared to Baseline

Model	Deviance	Level	Variance	Variance By Level
Model 0 (Null)	92931	Student	3625	29%
		School	6510	51%
		Country	2518	20%
Model 1 (Null + Stage 1 Effects)	92355	Student	3638	35%
		School	5205	50%
		Country	1585	15%
Model 2 (Null + Stage 1 Effects + Controls)	91720	Student	3539	40%
		School	4293	48%
		Country	1112	12%
Model 3 (Language Distance)	**91625***	**Student**	**3488**	**39%**
		School	**4267**	**48%**
		Country	**1079**	**12%**
Model 4 (HDI)	91718	Student	3539	40%
		School	4291	48%
		Country	1089	12%

Table G54 (cont'd)

Model	Deviance	Level	Variance	Variance By Level
Model 5 (GAIN)	91716*	Student	3540	40%
		School	4289	48%
		Country	1075	12%
Model 6 (FD Ratio)	91716*	Student	3538	40%
		School	4296	48%
		Country	1041	12%
Model 7 (Immigration Year HDI)	91672*	Student	3515	33%
		School	4169	39%
		Country	3047	28%
Model 8 (Immigration Year GAIN)	91695*	Student	3506	36%
		School	4276	43%
		Country	2079	21%
Model 9 (Immigration Year FD Ratio)	91715*	Student	3528	40%
		School	4319	48%
		Country	1058	12%

Note. Asterisk after the model name denotes the model deviance is statistically significantly different from the baseline Model 2. **Bold text** denotes models with covariates that were statistically significant within the model.

APPENDIX H: COMPLETE LIST OF IMMIGRATION PATHS

Table H1

Immigrant Students Paths from Origins to Destinations, Sorted by Country

#	Destination Country	Origin Country	Origin to Dest. Count	Dest. Cntry. Immig. Count
1	CAN	PHL	526	1705
2	CAN	USA	278	1705
3	CAN	CHN	253	1705
4	CAN	IND	171	1705
5	CAN	GBR	102	1705
6	CAN	PAK	100	1705
7	CAN	KOR	73	1705
8	CAN	FRA	66	1705
9	CAN	IRN	51	1705
10	CAN	SYRA	44	1705
11	CAN	ARE	41	1705
12	QAT	EGY	1035	1622
13	QAT	JOR	333	1622
14	QAT	YEM	254	1622
15	JOR	SYR	849	971
16	JOR	IRQ	87	971

Table H1 (cont'd)

#	Destination Country	Origin Country	Origin to Dest. Count	Dest. Cntry. Immig. Count
17	JOR	EGY	35	971
18	AUS	GBR	249	905
19	AUS	NZL	177	905
20	AUS	PHL	148	905
21	AUS	CHN	143	905
22	AUS	IND	143	905
23	AUS	VNM	27	905
24	AUS	GRC	10	905
25	AUS	ITA	8	905
26	NZL	GBR	161	548
27	NZL	ZAF	93	548
28	NZL	PHL	89	548
29	NZL	CHN	84	548
30	NZL	AUS	80	548
31	NZL	KOR	24	548
32	NZL	FJI	17	548
33	CHE	PRT	133	404

#	Destination Country	Origin Country	Origin to Dest. Count	Dest. Cntry. Immig. Count
34	CHE	ITA	95	404
35	CHE	DEU	89	404
36	CHE	FRA	32	404
37	CHE	ESP	24	404
38	CHE	TUR	13	404
39	CHE	AUT	10	404
40	CHE	ALB	8	404
41	BEL	NLD	103	247
42	BEL	DEU	72	247
43	BEL	FRA	66	247
44	BEL	TUR	8	247
45	AZE	RUS	154	201
46	AZE	GEO	42	201
47	AZE	TUR	5	201
48	ARG	BOL	91	191
49	ARG	PRY	67	191
50	ARG	BRA	15	191

Table H1 (cont'd)

#	Destination Country	Origin Country	Origin to Dest. Count	Dest. Cntry. Immig. Count
51	ARG	CHL	11	191
52	ARG	URY	7	191
53	ARG	NIC	99	215
54	CRI	NIC	171	191
55	CRI	COL	14	191
56	CRI	PAN	6	191
57	BIH	HRV	92	178
58	BIH	SRB	84	178
59	BIH	MNE	2	178
60	IRL	GBR	178	178
61	ISR	USA	84	156
62	ISR	ETH	43	156
63	ISR	FRA	29	156
64	GEO	RUS	133	148
65	GEO	AZE	8	148
66	GEO	ARM	7	148
67	AUT	DEU	71	127

Table H1 (cont'd)

#	Destination Country	Origin Country	Origin to Dest. Count	Dest. Cntry. Immig. Count
68	AUT	AFG	24	127
69	AUT	SYR	17	127
70	AUT	TUR	15	127
71	FIN	EST	43	122
72	FIN	RUS	31	122
73	FIN	CHN	17	122
74	FIN	SWE	15	122
75	FIN	IRQ	7	122
76	FIN	AFG	5	122
77	FIN	VNM	3	122
78	FIN	TUR	1	122
79	GBR	IRL	122	122
80	MDA	RUS	70	109
81	MDA	UKR	24	109
82	MDA	ROU	15	109
83	MNE	SRB	82	103
84	MNE	BIH	14	103

Table H1 (cont'd)

#	Destination Country	Origin Country	Origin to Dest. Count	Dest. Cntry. Immig. Count
85	MNE	ALB	6	103
86	MNE	HRV	1	103
87	PAN	VEN	55	102
88	PAN	COL	30	102
89	PAN	NIC	8	102
90	PAN	DOM	6	102
91	PAN	CHN	3	102
92	BLR	RUS	53	99
93	BLR	UKR	27	99
94	BLR	KAZ	15	99
95	BLR	POL	4	99
96	CZE	UKR	32	95
97	CZE	SVK	28	95
98	CZE	CHN	12	95
99	CZE	RUS	12	95
100	CZE	VNM	11	95
101	MEX	USA	79	77

Table H1 (cont'd)

#	Destination Country	Origin Country	Origin to Dest. Count	Dest. Cntry. Immig. Count
102	HRV	BIH	58	77
103	HRV	SRB	19	77
104	DEU	POL	31	75
105	DEU	ITA	10	75
106	DEU	TUR	9	75
107	DEU	HRV	8	75
108	DEU	BIH	7	75
109	DEU	GRC	5	75
110	DEU	SRB	4	75
111	DEU	MKD	1	75
112	DNK	SYR	19	67
113	DNK	ISL	9	67
114	DNK	AFG	7	67
115	DNK	IRQ	6	67
116	DNK	NOR	6	67
117	DNK	PAK	6	67
118	DNK	SWE	6	67

Table H1 (cont'd)

#	Destination Country	Origin Country	Origin to Dest. Count	Dest. Cntry. Immig. Count
119	DNK	TUR	6	67
120	DNK	FIN	2	67
121	MAR	ESP	22	64
122	MAR	FRA	21	64
123	MAR	ITA	18	64
124	MAR	DEU	3	64
125	SAU	JOR	25	63
126	SAU	KWT	13	63
127	SAU	QAT	13	63
128	SAU	USA	9	63
129	SAU	AUS	5	63
130	URY	ARG	33	63
131	URY	BRA	30	63
132	PRT	BRA	59	61
133	PRT	CHN	2	61
134	UKR	RUS	29	43
135	UKR	MDA	7	43

Table H1 (cont'd)

#	Destination Country	Origin Country	Origin to Dest. Count	Dest. Cntry. Immig. Count
136	UKR	BLR	4	43
137	UKR	KAZ	3	43
138	LVA	RUS	31	42
139	LVA	UKR	9	42
140	LVA	BLR	2	42
141	PHL	SAU	15	37
142	PHL	USA	10	37
143	PHL	CHN	8	37
144	PHL	ARE	4	37
145	PHL	SAU	15	42
146	SVK	CZE	21	33
147	SVK	HUN	12	33
148	IDN	MYS	17	33
149	IDN	NLD	9	33
150	IDN	AUS	3	33
151	IDN	SGP	3	33
152	KOR	USA	13	32

Table H1 (cont'd)

#	Destination Country	Origin Country	Origin to Dest. Count	Dest. Cntry. Immig. Count
153	KOR	JPN	8	32
154	KOR	CHN	7	32
155	KOR	PHL	2	32
156	KOR	RUS	2	32
157	NOR	SWE	17	29
158	NOR	DNK	12	29
159	TUR	DEU	18	26
160	TUR	RUS	5	26
161	TUR	BGR	2	26
162	TUR	NLD	1	26
163	MKD	ALB	10	20
164	MKD	BIH	5	20
165	MKD	SRB	5	20
166	DOM	ESP	8	18
167	DOM	HTI	6	18
168	DOM	USA	4	18
169	SVN	ITA	8	10

Table H1 (cont'd)

#	Destination Country	Origin Country	Origin to Dest. Count	Dest. Cntry. Immig. Count
170	SVN	HUN	2	10

APPENDIX I: IMMIGRATION COUNTS AND PERCENTAGES BY COUNTRY

Further analysis of the asymmetric immigration paths was conducted for each country. This analysis shows how many immigrant students are found in each destination country (see Table I1). The first column is the row number. The second column indicates the country, the third column shows how many immigrant students ended up in the aforementioned country. The fourth column shows the percentage of said country's PISA sample with an immigrant background, among immigrant and native students. An illustrative example result is that Canada (i.e., CAN) had 1705 students with an immigrant background in their PISA 2018 sample. Furthermore, Canada's PISA 2018 sample was 20.70% immigrant students.

Table I1

Immigration Counts and Percentages by Destination Country

#	Country	Immigrant Count	Immigrant % Population
1	CAN	1705	20.70
2	QAT	1622	40.26
3	JOR	971	17.22
4	AUS	905	21.43
5	NZL	548	22.57
6	CHE	404	30.55
7	BEL	247	13.06
8	AZE	201	4.87

Table I1 (cont'd)

#	Country	Immigrant Count	Immigrant % Population
9	ARG	191	4.74
10	CRI	191	9.44
11	BIH	178	2.50
12	IRL	178	10.25
13	ISR	156	13.94
14	GEO	148	1.03
15	AUT	127	16.60
16	FIN	122	4.57
17	GBR	122	7.99
18	MDA	109	1.23
19	MNE	103	5.16
20	PAN	102	4.66
21	BLR	99	3.78
22	GRC	98	10.25
23	CZE	95	3.30
24	MEX	79	0.78
25	HRV	77	8.78
26	DEU	75	18.71

Table I1 (cont'd)

#	Country	Immigrant Count	Immigrant % Population
27	DNK	67	19.35
28	MAR	64	0.71
29	SAU	63	7.35
30	URY	63	0.97
31	PRT	61	4.78
32	UKR	43	1.88
33	LVA	42	4.31
34	PHL	37	0.64
35	SVK	33	0.76
36	IDN	32	0.24
37	KOR	32	0.09
38	NOR	29	7.01
39	TUR	26	0.59
40	MKD	20	1.10
41	DOM	18	2.44
42	SVN	10	4.38

Table I2 shows the unique country-to-country pairings of immigrant students within the sample. analysis of the asymmetric immigration paths was conducted for each country. The first column is the row number. The second column indicates the destination country, the third column shows the origin country. The fourth column is the count of cases for that particular country pair. An illustrative example is that Qatar to Egypt had 1035 unique cases.

Table I2

Immigrant Students Paths from Origins to Destinations, Sorted by Count

#	Destination Country	Origin Country	Origin to Destination Count
1	QAT	EGY	1035
2	JOR	SYR	849
3	CAN	PHL	526
4	QAT	JOR	333
5	CAN	USA	278
6	QAT	YEM	254
7	CAN	CHN	253
8	AUS	GBR	249
9	IRL	GBR	178
10	AUS	NZL	177
11	CAN	IND	171
12	CRI	NIC	171

Table I2 (cont'd)

#	Destination Country	Origin Country	Origin to Destination Count
13	NZL	GBR	161
14	AZE	RUS	154
15	AUS	PHL	148
16	AUS	CHN	143
17	AUS	IND	143
18	CHE	PRT	133
19	GEO	RUS	133
20	GBR	IRL	122
21	BEL	NLD	103
22	CAN	GBR	102
23	CAN	PAK	100
24	GRC	ALB	98
25	CHE	ITA	95
26	NZL	ZAF	93
27	BIH	HRV	92
28	ARG	BOL	91
29	CHE	DEU	89
30	NZL	PHL	89

Table I2 (cont'd)

#	Destination Country	Origin Country	Origin to Destination Count
31	JOR	IRQ	87
32	BIH	SRB	84
33	ISR	USA	84
34	NZL	CHN	84
35	MNE	SRB	82
36	NZL	AUS	80
37	MEX	USA	79
38	CAN	KOR	73
39	BEL	DEU	72
40	AUT	DEU	71
41	MDA	RUS	70
42	ARG	PRY	67
43	CAN	FRA	66
44	BEL	FRA	64
45	PRT	BRA	59
46	HRV	BIH	58
47	PAN	VEN	55
48	BLR	RUS	53

Table I2 (cont'd)

#	Destination Country	Origin Country	Origin to Destination Count
49	CAN	IRN	51
50	CAN	SYR	44
51	FIN	EST	43
52	ISR	ETH	43
53	AZE	GEO	42
54	CAN	ARE	41
55	JOR	EGY	35
56	URY	ARG	33
57	CHE	FRA	32
58	CZE	UKR	32
59	DEU	POL	31
60	FIN	RUS	31
61	LVA	RUS	31
62	PAN	COL	30
63	URY	BRA	30
64	ISR	FRA	29
65	UKR	RUS	29
66	CZE	SVK	28

Table I2 (cont'd)

#	Destination Country	Origin Country	Origin to Destination Count
67	AUS	VNM	27
68	BLR	UKR	27
69	SAU	JOR	25
70	AUT	AFG	24
71	CHE	ESP	24
72	MDA	UKR	24
73	NZL	KOR	24
74	MAR	ESP	22
75	MAR	FRA	21
76	SVK	CZE	21
77	DNK	SYR	19
78	HRV	SRB	19
79	MAR	ITA	18
80	TUR	DEU	18
81	AUT	SYR	17
82	FIN	CHN	17
83	IDN	MYS	17
84	NOR	SWE	17

Table I2 (cont'd)

#	Destination Country	Origin Country	Origin to Destination Count
85	NZL	FJI	17
86	ARG	BRA	15
87	AUT	TUR	15
88	BLR	KAZ	15
89	FIN	SWE	15
90	MDA	ROU	15
91	PHL	SAU	15
92	CRI	COL	14
93	MNE	BIH	14
94	CHE	TUR	13
95	KOR	USA	13
96	CZE	CHN	12
97	CZE	RUS	12
98	NOR	DNK	12
99	SAU	KWT	12
100	SAU	QAT	12
101	SVK	HUN	12
102	ARG	CHL	11

Table I2 (cont'd)

#	Destination Country	Origin Country	Origin to Destination Count
103	CZE	VNM	11
104	AUS	GRC	10
105	CHE	AUT	10
106	DEU	ITA	10
107	MKD	ALB	10
108	PHL	USA	10
109	DEU	TUR	9
110	DNK	ISL	9
111	IDN	NLD	9
112	LVA	UKR	9
113	SAU	USA	9
114	AUS	ITA	8
115	BEL	TUR	8
116	CHE	ALB	8
117	DEU	HRV	8
118	DOM	ESP	8
119	GEO	AZE	8
120	KOR	JPN	8

#	Destination Country	Origin Country	Origin to Destination Count
121	PAN	NIC	8
122	PHL	CHN	8
123	SVN	ITA	8
124	ARG	URY	7
125	DEU	BIH	7
126	DNK	AFG	7
127	FIN	IRQ	7
128	GEO	ARM	7
129	KOR	CHN	7
130	UKR	MDA	7
131	CRI	PAN	6
132	DNK	IRQ	6
133	DNK	NOR	6
134	DNK	PAK	6
135	DNK	SWE	6
136	DNK	TUR	6
137	DOM	HTI	6
138	MNE	ALB	6

Table I2 (cont'd)

#	Destination Country	Origin Country	Origin to Destination Count
139	PAN	DOM	6
140	AZE	TUR	5
141	DEU	GRC	5
142	FIN	AFG	5
143	MKD	BIH	5
144	MKD	SRB	5
145	SAU	AUS	5
146	TUR	RUS	5
147	BLR	POL	4
148	DEU	SRB	4
149	DOM	USA	4
150	PHL	ARE	4
151	UKR	BLR	4
152	FIN	VNM	3
153	IDN	AUS	3
154	IDN	SGP	3
155	MAR	DEU	3
156	PAN	CHN	3

Table I2 (cont'd)

#	Destination Country	Origin Country	Origin to Destination Count
157	UKR	KAZ	3
158	BIH	MNE	2
159	DNK	FIN	2
160	KOR	PHL	2
161	KOR	RUS	2
162	LVA	BLR	2
163	PRT	CHN	2
164	SVN	HUN	2
165	TUR	BGR	2
166	DEU	MKD	1
167	FIN	TUR	1
168	MNE	HRV	1
169	TUR	NLD	1