Chapter 5

Language barriers and the resilience of students with an immigrant background

Immigrant students face multiple sources of disadvantage that affect their academic performance and their general well-being. Fluency in the language spoken in the host country is one of these source factors. Language barriers can also amplify the effects of other sources of disadvantage, such as having migrated after the age of 12, lack of parental support, studying in a disadvantaged school or attending a school with a poor disciplinary climate. This chapter examines language as a risk factor when considering the academic, social, emotional and motivational resilience of immigrant students.

Notes regarding Cyprus

Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Note regarding data from Israel

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

What the data tell us

- The language spoken at home strongly affects the likelihood that immigrant students will be academically resilient. On average across OECD countries, immigrant students who do not speak the language of assessment at home are around eight percentage points less likely to be academically resilient than native-speaking immigrant students (nine percentage points less likely across EU countries).
- On average across OECD countries, native-speaking immigrant students with at least one foreign-born parent are two percentage points less likely than native students to attain baseline levels of proficiency in the core PISA subjects, while non-native-speaking immigrant students are about 17 percentage points less likely to do so (EU figures 3 and 18 percentage points).
- Immigrant students who are non-native speakers are five percentage points less likely than those who are native speakers to report a sense of belonging at school, on average across OECD countries (six percentage points less likely across EU countries).
- The greater the linguistic distance between the language spoken at home and the language of instruction, the less likely a student will attain baseline academic proficiency and report a sense of belonging at school.

Research indicates that fluency in the host-country language is one of the most important determinants of social and economic integration of immigrant students (OECD, 2006). Since humans communicate first and foremost by speaking, writing and reading, learning the language spoken in the host country has an immediate impact on immigrants' lives and on their integration. Investing in acquiring the host-country language has a positive impact on almost all facets of life but especially so on the integration of immigrant children in education settings (Van Tubergen, Maas and Flap, 2004; Van Tubergen and Kalmijn, 2005).

In the economic domain, language fluency is positively associated with higher productivity and wages because it facilitates the transfer and adaptation of skills in the job market (Chiswick and Miller, 1995; Hayfron, 2001; Dustmann and Fabbri, 2003; Dustmann and Soest, 2001). Likewise, immigrants with good language skills are more likely to be better at searching for work and landing jobs that match their skills and qualifications (Leslie and Lindley, 2001; Frijters, Shields and Price, 2005). It is also easier for immigrants who are fluent in the host-country language to acquire additional education, training and experience in the labour market.

Figure 5.1 ▪ **The role of the language spoken at home in the resilience process**

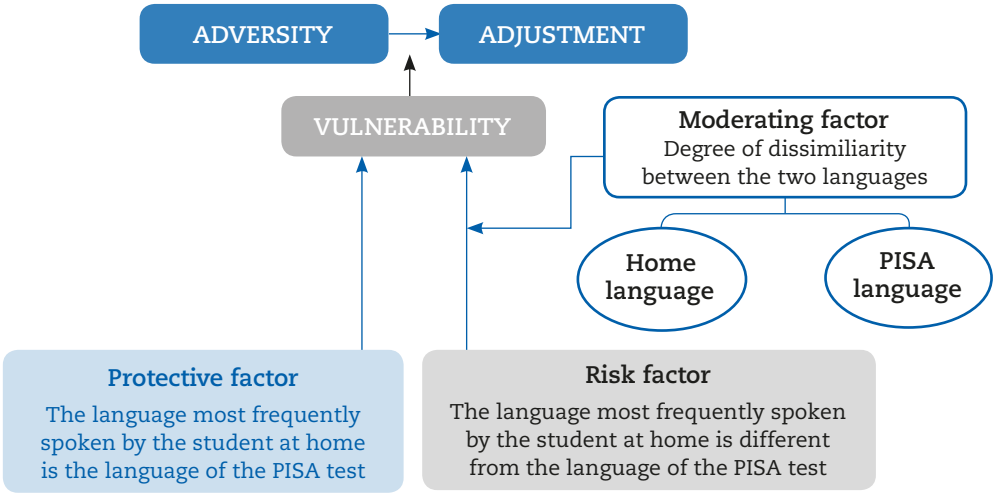


Table 5.1 ▪ Snapshot of language barriers to attaining baseline academic proficiency

Countries/economices with values **above** the OECD average
Countries/economices with values not significantly different from the OECD average
Countries/economices with values **below** the OECD average

	Share of immigrant students who are non-native speakers	Relative risk of native-speaking immigrant students not being academically resilient (compared to native-speaking native students)	Relative risk of non-native speaking immigrant students not being academically resilient (compared to non-native-speaking native students)	Share of immigrant students with at least one native-born parent who are non-native speakers	Relative risk of native-speaking immigrant students with at least one native-born parent not being academically resilient (compared to native-speaking native students)	Relative risk of non-native-speaking immigrant students with at least one native-born parent not being academically resilient (compared to non-native-speaking native students)
OECD average	**49.18**	**1.52**	**1.12**	**12.74**	**1.05**	**0.99**
EU average	**48.50**	**1.59**	**1.20**	**17.79**	**1.11**	**1.02**
Lebanon	97.87	c	1.05	97.38	c	0.83
Iceland	83.25	c	c	9.58	1.04	c
Finland	76.95	2.56	1.93	12.11	1.15	1.51
Luxembourg	76.54	c	2.38	83.82	c	1.51
Austria	74.25	1.62	1.34	22.86	1.05	1.17
Slovenia	73.58	1.84	0.89	12.02	1.19	0.92
Singapore	70.02	0.73	0.4	52.65	0.88	0.67
Sweden	68.09	1.9	1.66	13.52	1.05	1.2
United States	66.93	1.2	0.8	13.44	1.03	0.84
Czech Republic	65.82	1.71	0.68	10.94	1.13	0.69
Switzerland	65.1	1.83	1.53	17.7	1.33	1.2
Japan	62.04	c	c	6.3	1.31	c
Malta	61.72	1.75	0.93	72.19	1.56	0.97
Italy	59.13	1.48	1.21	15.86	1	1.06
Norway	58.74	1.79	0.97	8.08	1.09	0.81
Germany	54.33	1.99	0.98	15.89	1.51	1.03
Cyprus*	53.18	0.92	1.58	25.29	0.92	1.1
Belgium	52.97	2.03	1.54	23.78	1.33	1.16
Spain	52.27	1.72	1.96	23.66	0.91	1.13
Bulgaria	52.11	c	0.82	16.9	1.15	0.92
Slovak Republic	51.68	1.8	0.87	10.02	1.18	0.89
Denmark	50.54	2.37	1.36	7.71	1.02	1.07
Canada	49.52	0.89	0.72	8.97	0.96	0.83
Ireland	49.46	1.02	c	0.91	0.89	c
United Arab Emirates	48.63	0.58	0.52	23.25	0.9	0.86
Qatar	47.84	0.61	0.56	30.06	0.9	0.95
Netherlands	47.73	1.67	0.82	10.69	1.19	0.71
United Kingdom	44.04	1.2	0.93	5.77	0.98	1.21
Israel	42.44	1.06	0.72	8.97	0.7	0.69
France	41.41	1.86	0.94	9.97	1.1	0.98
New Zealand	41.07	0.94	0.55	3.45	0.82	c
Australia	38.38	0.84	0.46	4.49	0.84	0.71
Greece	35.58	1.46	0.85	6.33	1.01	0.54
Tunisia	33.59	1.15	c	20.36	0.99	1.08
Turkey	30.74	1.09	c	5.15	0.68	c
Uruguay	28.6	c	c	12.67	1.02	0.98
Dominican Republic	27.96	1	c	10.21	0.98	c
Albania	27.91	c	c	7.63	1.03	c
Lithuania	26.72	0.97	0.72	16.14	1.12	0.74
FYROM	26.38	1.17	c	14	0.98	0.87
Thailand	25.99	1.1	c	9.22	0.99	c
Latvia	25.59	1.15	1.3	16.34	1.09	1.05
Portugal	25.4	1.29	c	3.09	0.82	c
B-S-J-G (China)	21.96	c	c	2.48	2.93	c
Mexico	20.93	1.54	c	8.87	1.06	c
Hungary	20.1	0.84	0.71	6.91	0.75	c
Trinidad and Tobago	18.78	1.24	c	3.04	0.95	c
Brazil	15.02	1.3	c	8.36	1.03	1.26
Russia	14.96	1.07	1.42	1.1	0.96	c
Peru	14.8	c	c	15.42	1.04	c
Macao (China)	14.13	0.53	0.84	20.44	0.77	0.8
Moldova	14.13	0.9	c	14.38	0.87	1.07
Estonia	13.82	1.58	1.09	17.91	1.36	0.76
Georgia	13.79	0.96	c	13.22	0.93	0.91
Montenegro	13.38	0.87	0.76	3.29	0.86	0.9
Algeria	13.2	1.04	c	21.38	0.98	c
CABA (Argentina)	12.43	1.92	c	5.02	1.28	c
Colombia	8.39	1.27	c	4.25	1.13	c
Croatia	7.92	1.23	1.05	4.54	0.96	1.04
Hong Kong (China)	7.58	1.14	c	3.11	1.19	c
Jordan	6.42	0.91	1.09	7.57	0.85	0.97
Costa Rica	4.69	1.24	c	5.32	1.05	c
Chile	4.46	1.32	c	6.45	0.88	c
Kosovo	3.29	1.06	c	2.17	0.93	c
Korea	c	c	c	1.66	1	c
Poland	c	c	c	12.84	1.13	c
Indonesia	c	c	c	45.73	c	c
Romania	c	c	c	20.16	1.17	c
Chinese Taipei	c	c	c	2.23	1.14	c
Viet Nam	c	c	c	5.3	0.92	c

* See note at the beginning of this Chapter.

Notes: Native-speaking students are students who speak most frequently at home the language of the PISA assessment.
Non-native-speaking students are students who reported that the language they most frequently speak at home is different from the language of the PISA assessment.
Academically resilient students are students with an immigrant background who attained at least proficiency Level 2 in all three core PISA subjects: science, reading and mathematics.

Source: OECD, PISA 2015 Database, Table 5.4.

StatLink http://dx.doi.org/10.1787/888933681692

Table 5.2 ▪ Snapshot of language barriers to feeling a sense of belonging at school

Countries/economies with values **above** the OECD average
Countries/economies with values not significantly different from the OECD average
Countries/economies with values **below** the OECD average

	Relative risk of native-speaking immigrant students not being socially resilient (compared to native-speaking native students) (sense of belonging at school)	Relative risk of non-native-speaking immigrant students not being socially resilient (compared to non-native-speaking native students) (sense of belonging at school)	Relative risk of native-speaking immigrant students with at least one native-born parent not being socially resilient (compared to native-speaking native students) (sense of belonging at school)	Relative risk of non-native-speaking immigrant students with at least one native-born parent not being socially resilient (compared to non-native-speaking native students) (sense of belonging at school)
OECD average	**1.11**	**1.03**	**1.08**	**1.04**
EU average	**1.15**	**1.08**	**1.10**	**1.06**
Australia	0.74	0.65	0.93	0.95
Norway	0.74	0.63	1.06	0.94
Lithuania	0.77	0.96	1.03	0.88
Luxembourg	0.79	1.56	0.86	1.21
New Zealand	0.8	0.6	0.96	c
United Kingdom	0.8	1.23	0.97	0.88
Finland	0.84	0.81	1.11	0.99
Qatar	0.85	0.65	1.01	0.94
Canada	0.89	0.76	1.06	0.96
Hungary	0.91	1.04	0.97	c
Singapore	0.97	0.84	1.11	0.94
Georgia	0.98	c	1.28	1.3
Chile	0.99	c	1	c
Austria	0.99	1.16	0.97	1.26
Costa Rica	1.03	c	1.14	c
United Arab Emirates	1.03	0.65	1.08	0.88
Macao (China)	1.05	1.04	1.13	0.87
Hong Kong (China)	1.05	1.34	1.02	1
Russia	1.1	1.2	1.07	c
France	1.11	0.89	1.02	1.02
United States	1.11	1.16	1.1	1.11
Netherlands	1.11	1.2	1.12	1.31
Croatia	1.12	0.78	1.02	0.87
Germany	1.13	1	1.14	1.08
Italy	1.13	1.14	1.11	0.93
Moldova	1.16	c	1.15	0.87
Slovenia	1.17	0.68	1.09	0.88
Sweden	1.18	0.86	1.2	0.88
Jordan	1.18	1.18	1.08	1.01
Turkey	1.19	c	1.07	c
Greece	1.2	0.96	1.06	0.97
Czech Republic	1.2	1	0.95	1
Montenegro	1.21	1.23	1.11	1.55
Ireland	1.24	m	1.04	m
Portugal	1.25	0.58	1	0.56
Malta	1.27	1.41	1.16	1.04
Belgium	1.27	1.22	1.25	1.31
Denmark	1.29	0.85	1.07	0.97
Switzerland	1.34	0.96	1.17	0.86
Thailand	1.34	c	1.2	c
Latvia	1.38	1.46	1.16	1.46
Estonia	1.38	1.16	1.46	0.89
Dominican Republic	1.41	c	1.09	c
Spain	1.45	1.49	1.21	1.37
Tunisia	1.47	c	1.3	1.23
Trinidad and Tobago	1.53	c	1.21	c
Cyprus*	1.55	1.09	1.32	1.11
Colombia	1.6	c	1.36	c
Mexico	1.61	c	1.27	c
Brazil	1.65	c	1.36	c
CABA (Argentina)	1.65	0.84	1.43	c
Kosovo	1.83	c	0.99	c
FYROM	2.94	c	1.21	0.69
Iceland	c	1.19	0.84	0.79
Japan	c	c	1.05	c
Korea	c	c	0.99	c
Poland	c	c	1.25	c
Slovak Republic	c	1.65	1.04	1.44
Albania	c	c	1.36	c
Algeria	c	c	1.04	c
B-S-J-G (China)	c	c	1.26	c
Bulgaria	c	c	1.13	1.31
Indonesia	c	c	c	c
Lebanon	c	1.16	c	0.95
Peru	c	c	1.58	c
Romania	c	c	1.06	c
Chinese Taipei	c	c	1.14	c
Uruguay	c	c	1.15	1.33
Viet Nam	c	c	1.47	c

* See note at the beginning of this Chapter.

Notes: Only countries/economies with valid data for at least one outcome are presented.
Native-speaking students are students who speak most frequently at home the language of the PISA assessment.
Non-native-speaking students are students who reported that the language they most frequently speak at home is different from the language of the PISA assessment.
Socially resilient students are students with an immigrant background who reported that they "agree" or "strongly agree" with the statement "I feel like I belong at school" and "disagree" or "strongly disagree" with the statement "I feel like an outsider at school".
Source: OECD, PISA 2015 Database.
StatLink http://dx.doi.org/10.1787/888933681711

The positive effects of language proficiency go beyond the labour market, and encompass health, marriage, social integration and political participation (Isphording, 2015; Clarke and Isphording, 2017). Immigrants who are fluent in the host-community language can interact more frequently and intensively with local populations and have greater access to information (Espenshade and Calhoun, 1993; Gordon, 1964). Increased intercultural contact makes it easier for immigrants to learn the values, norms and traditions in the host country and eases the way towards integration (Chiswick and Miller, 1996).

Since the 1990s, language training has been a key policy priority in many countries to ensure that immigrants are able to contribute to the economic and social life of their communities and are, in turn, also to benefit from them (Joppke, 2007). Civic integration policies introduced in the Netherlands were among the first in Europe to emphasise knowledge of host country language and culture. Shortly after, language policies were implemented by other countries, including Austria, Denmark, Finland, France and Germany. Eventually, the focus on language acquisition also became one of the principles of the EU statement on integration policy. The fourth principle states: "Basic knowledge of the host society's language, history, and institutions is indispensable to integration" (Council of the European Union 2004, 20). The obligatory character of civic integration policies and their strong focus on language acquisition exemplify the importance given to language for integration.

Just as fluency in the language spoken in the host community is important for adults because it ensures their economic and social integration, language fluency among school-aged children ensures that they are able to make the most of the learning opportunities offered by schools (Boykin, Tyler and Miller, 2005; Dustmann, Machin and Schonberg, 2010; Geay, McNally and Telhaj, 2013). While the effect of language use at home on the academic outcomes of immigrants has been explored in depth both in the empirical and theoretical literature (OECD, 2015b; OECD, 2010; OECD, 2006), less is known about the role language fluency plays in shaping the socio-emotional and motivational resilience of immigrant students.

Language fluency enables children with an immigrant background to participate actively in the social life of their school, and develop a sense of belonging at their school community and beyond (Coll and Magnuson, 1997; Zhou and Xiong, 2005; Dawson and Williams, 2008). Language facilitates the socialisation of children with an immigrant background in their new environment and supports their acculturation in the destination country.

Children with an immigrant background with language difficulties are found to be more likely to be bullied, discriminated against and are more likely to suffer emotional problems, such as depression and low self-esteem (Gil, Vega and Dimas, 1994; Padilla and Perez, 2003; Romero and Roberts, 2003; Smart and Smart, 1995). The sooner children become fluent in the host-country language, the more they can benefit from new opportunities. Thus, in many countries, schools emphasise language acquisition for children with an immigrant background.

Language barriers can also make it difficult for immigrant parents to help their children integrate into their new community. For instance, parents who are not fluent in the host-country language might be particularly at risk of being out of the labour force or experience economic hardship. This may hinder them from providing the material resources or the intangible assistance, such as supporting them in school and course selection and encouraging good study strategies, that could help in their children's education (Bermudez, 1994; Moles, 1993).

Unfortunately large-scale international assessments have limited information on the mother tongue of immigrant students or the variety of languages students speak within the family, the choices such students and their families make about the use of their mother tongue or the language of instruction within the home and the reasons behind such choices, the level of proficiency in different languages students and their families possess and the context in which they use specific languages. Although language is crucial to promote immigrant students' integration, the classification of language groups developed in this chapter is based on limited set of information contained in the PISA dataset: the language of the PISA test and the language that the student reports speaking most frequently at home.

The language of assessment generally reflects the language that is spoken in the country that conducted the PISA test and in which students are taught. However, in some countries and contexts multiple language communities coexist and students are taught in more than one language.

The PISA background questionnaire asked students to report which language they speak most frequently at home. In this report ton-native speakers are students who reported that the language they speak most frequently at home is different from the language of the PISA test, while native speakers are students who reported that the language they speak most frequently at home is the same as that of the PISA test. This classification may hide important differences: some students may be bilingual and be native speakers in multiple languages, other students may speak at home the language of instruction even if they are non-native speakers, because they and their parents may feel this could help them gain proficiency at a faster pace. In this chapter, native students are defined as those without an immigrant background who speak most frequently at home the language of the PISA test.[1, 2]

Individual characteristics and use of the host-country language at home

How recently a child (or his or her family) immigrated, their socio-economic status and their country of origin have all been shown to be correlated with the use of host-country language at home. The effects of these individual characteristics are understood better when situated within the theoretical framework developed by Chiswick and Miller (1995). Their framework classifies the determinants of language acquisition into three groups: factors that affect immigrants' exposure to the language of the host country, factors that shape immigrants' ability to become fluent in a new language, and factors that influence the efficiency with which immigrants become fluent.

Immigrant students' exposure to the host-country language is determined by the amount of time they have spent in the host country, the number of interactions that occur per unit of time, on average, and the efficacy of those interactions. Second-generation immigrant children are born and raised in the host country. Their schooling has taken place only within the host country and they have been exposed to the destination-country language for longer than first-generation immigrant students, especially if they had also attended pre-primary school. It is thus easier for second-generation immigrant students than for first-generation immigrant students to be fluent in the destination-country language. Moreover, immigrants who have lived in the host country for longer will have been more exposed to the host-country language and are more likely to be fluent in it (Isphording and Otten, 2013).

Immigrants' age at arrival in the host country matters too. Immigrant students who had arrived as young children might be able to acquire the host-country language with much greater ease (Newport, 2002). In fact, supporters of the Critical Period Hypothesis argue that the age of 12 marks an important threshold, after which the efficiency with which people acquire foreign-language skills decreases markedly. Thus, first-generation immigrants (compared to second-generation immigrants), immigrants who have been in the country for a shorter period, and immigrants who had arrived after the age of 12 might be at a double disadvantage when it comes to overcoming language barriers.

Research has consistently shown that parents' educational background is one of the most significant determinants of immigrant children's academic and well-being outcomes (Capps, 2005; Bilgili et al., 2015). Immigrant children of highly educated parents are also more likely than their peers with low-educated parents to become proficient more quickly in a new language. Highly educated people who are also proficient in their native language might be better able to understand what is required to master a new language, and how to seek and access support to do so.

The official language of the country of origin is also an important determinant of language use at home. Countries of origin and destination might share an historical link, such as through colonialism, whereby immigrants from former colonies might already speak the language of the host country in addition to a second language or local dialect. Language, itself, might be the bond that links a country of origin to a destination country. For example, there are strong migration flows among Spanish-speaking countries. In both of these cases, an immigrant child does not have to learn a new language when he or she arrives in the host country.

For many children with an immigrant background, however, learning the host-country's language is an enormous challenge; but difficulties might be reduced if the language in the country of origin has similar roots as the host-country language. No research to date has explored the relationship between language similarity, or proximity, and migrant students' academic, socio-emotional and motivational resilience. When the difference between two languages is larger, it might be more difficult to learn the host-country language, which could have implications for the resilience of students with an immigrant background.

Immigrant background and languages spoken at home: An overview

PISA reveals that in 2015, on average across OECD countries, 12% of students did not speak the language of assessment as their main language at home (15% on average across EU countries). However, there were considerable differences across countries. For example, in Luxembourg, 84% of students did not speak the language of assessment at home, while in Japan, only 1% of students did not speak the language of assessment at home.

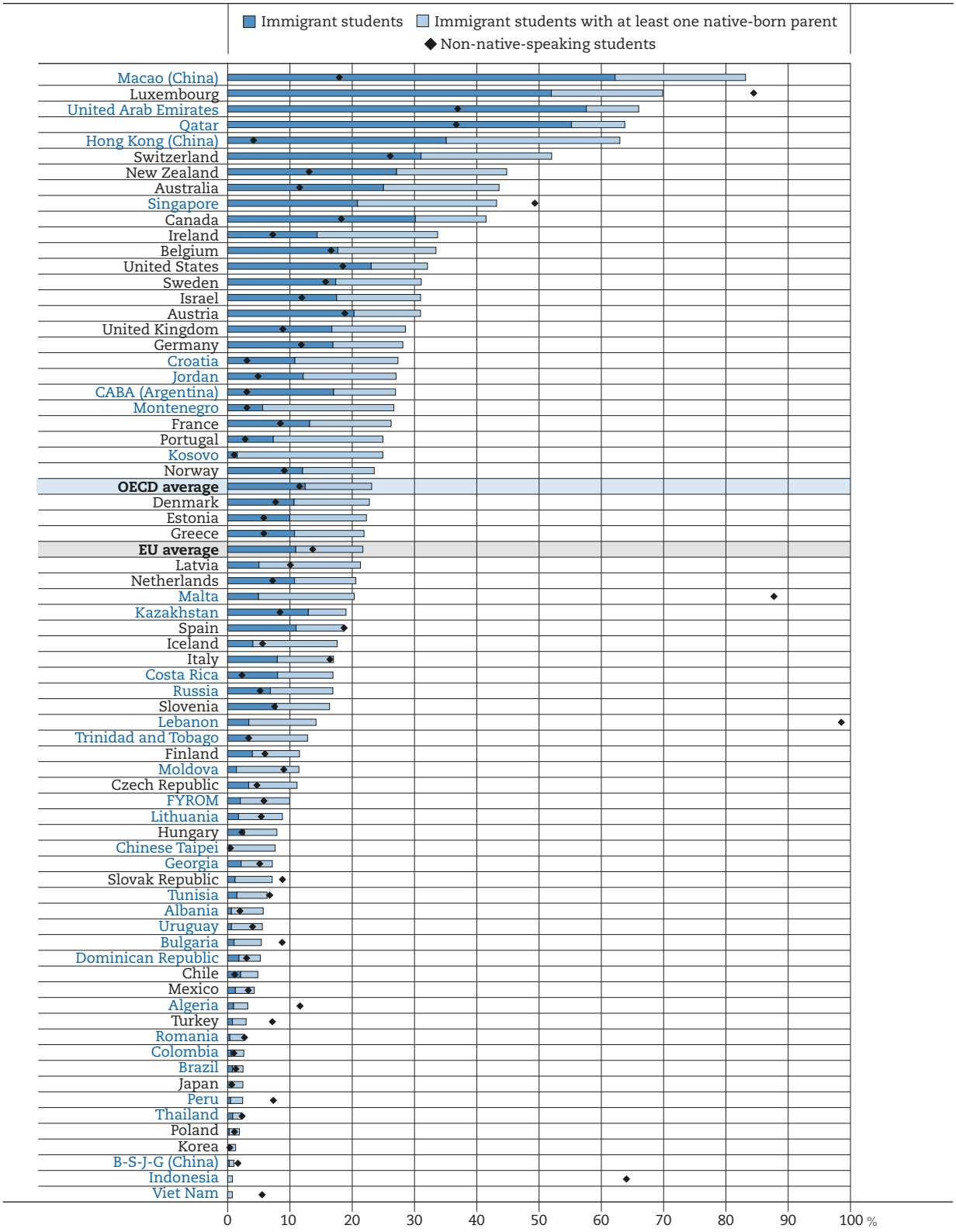
PISA data also reveal that, while there is an association between having an immigrant background and not speaking the language of assessment at home, one doesn't necessarily imply the other. For example, although the percentage of immigrant students and the percentage of students who do not speak the language of assessment at home were practically the same in 2015 at the OECD average level (12% and 15%, respectively, at the EU average level), the two groups do not overlap perfectly. In fact, on average across OECD countries, 49% of immigrant students and 6% of native students did not speak the language of assessment at home (49% and 10%, respectively, across EU countries). Among students who did not speak the language of assessment at home, 68% were students with an immigrant background and 32% were native students (61% and 39%, respectively, across EU countries).

Figure 5.2 shows the percentages of immigrant students, native-born students of mixed heritage, and students who do not speak the language of assessment at home in the countries and economies that participated in PISA 2015. In the majority of them, there were more immigrant students and native students of mixed heritage than students who speak at home a language that is different from the language of instruction. In a number of countries, including Australia, Canada, New Zealand and the United Kingdom, the share of immigrant students was larger than the share of students who speak a different language at home. These are countries where most immigrants come from territories where the host language is spoken – predominantly former colonies of the British Empire. By contrast, in Indonesia, Lebanon, Luxembourg, Malta and Singapore, the share of students who do not speak the language of assessment was larger than the share of immigrant students. These countries are home to established language minorities.

PISA data reveal that in 2015 speaking a language at home that is different from the language of the assessment was more frequent among immigrant students. Figure 5.3a shows the percentage of native students and of first- and second-generation immigrant students participating in PISA 2015 who reported not speaking the language of assessment at home. Significant differences between immigrant students (first- and second-generation combined) and native students are reported next to the name of each country or economy. In all countries and economies shown except Luxembourg, Macao (China) and Malta, the percentage of students who reported not speaking the language of assessment at home was greater among immigrant students than among native students.

The figure also shows that in most countries it was more common for first-generation immigrant students than for second-generation immigrant students to speak a language at home that is different from the language of assessment. On average across OECD countries, 6% of native students (10% across EU countries), 60% of first-generation immigrant students (60% across EU countries), and 41% of second-generation immigrant students (40% across EU countries) reported not speaking the language of assessment at home. In Austria, Finland, Iceland and Slovenia, more than 70% of immigrant students, but only 5% or less of native students, reported speaking a language at home that is different from the language of assessment. Luxembourg, Macao (China) and Malta are the only countries and economies where it was more common for native students to speak a language at home that is different from the language of instruction.

Figure 5.2 ■ **Percentage of students with an immigrant background and non-native-speaking students**



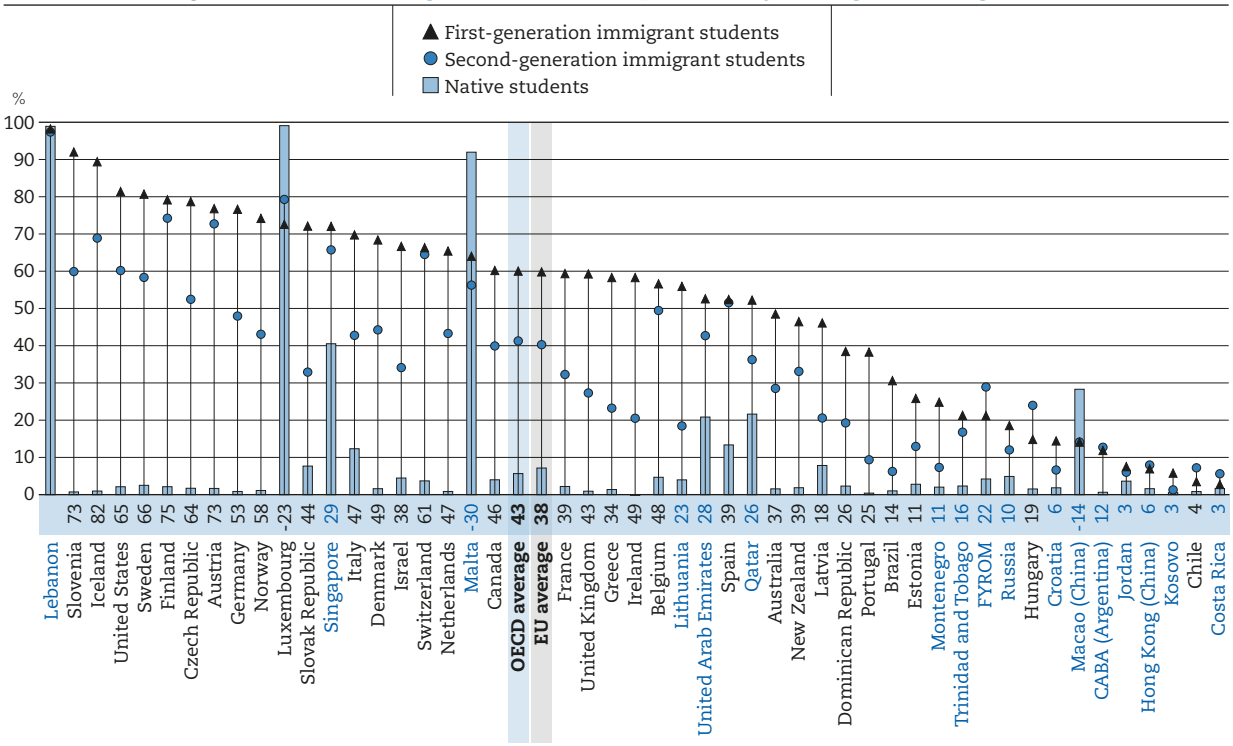
Notes: Non-native-speaking students are students who reported that the language they most frequently speak at home is different from the language of the PISA assessment.

Countries and economies are ranked in descending order of the percentage of students with an immigrant background.

Source: OECD, PISA 2015 Database, Table 5.3.

StatLink http://dx.doi.org/10.1787/888933681464

Figure 5.3a ▪ **Percentage of non-native speakers, by immigrant background**



Notes: Non-native-speakers are students who reported that the language they most frequently speak at home is different from the language of the PISA assessment.

Only countries with valid values for first- and second-generation immigrant students are shown.

Statistically significant differences in the percentage of non-native-speaking students among immigrant and native students are reported next to country/economy names.

Countries and economies are ranked in descending order of the percentage of first-generation immigrant students who do not speak the language of assessment at home.

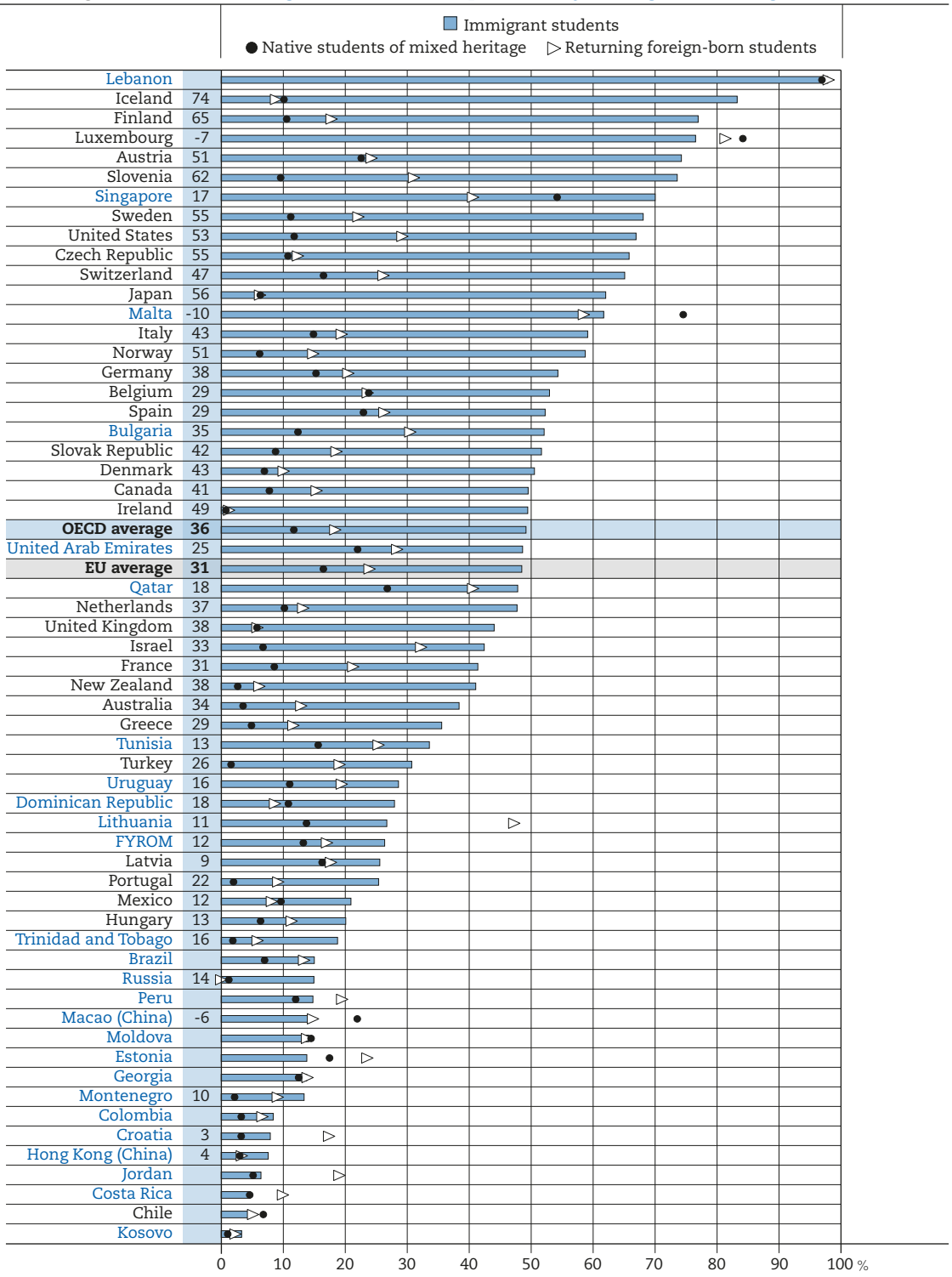
Source: OECD, PISA 2015 Database, Table 5.4.

StatLink http://dx.doi.org/10.1787/888933681483

Figure 5.3b shows the percentage of returning foreign-born students (that is, foreign-born students who have at least one native-born parent), native students of mixed heritage (native students with one native-born and one foreign-born parent) and immigrant students (a category which comprises both native and foreign-born students who have foreign born parents) who do not speak the language of assessment at home. In all countries and economies except Brazil, Chile, Colombia, Costa Rica, Estonia, Georgia, Jordan, Kosovo, Lebanon, Luxembourg, Malta, Moldova and Peru, immigrant students with two foreign-born parents were more likely to speak a language at home that is different from the language of assessment compared to returning foreign-born students and native students of mixed-heritage (students with an immigrant background who have at least one native-born parent). In Finland, Iceland and Slovenia, more than 70% of immigrant students with two foreign-born parents reported not speaking the language of assessment at home while less than 15% of other students with an immigrant background who have at least one native-born parent so reported. Luxembourg, Macao (China) and Malta were the only countries and economies where students with an immigrant background who have at least one native-born parent were more likely to be non-native speakers than the children of foreign-born parents.

PISA data indicate that among all groups of students with an immigrant background, first-generation students were the least likely to speak the language of instruction at home in 2015, particularly if they arrived in the country in which they sat the PISA test at or after the age of 12 (Figure 5.3c). On average across OECD the share of non-native speakers among late arrivals was 13 percentage points larger than that of non-native speakers among immigrant students who had arrived before the age of 12 (early arrivals) (14% across EU countries). In the Czech Republic, Finland, Slovenia and Sweden, over 90% of late arrivals reported that they did not speak the language of assessment at home.

Figure 5.3b ■ **Percentage of non-native speakers, by immigrant heritage**



Notes: Non-native-speakers are students who reported that the language they most frequently speak at home is different from the language of the PISA assessment.

Only countries with valid values for returning foreign-born students and native students of mixed heritage are shown.

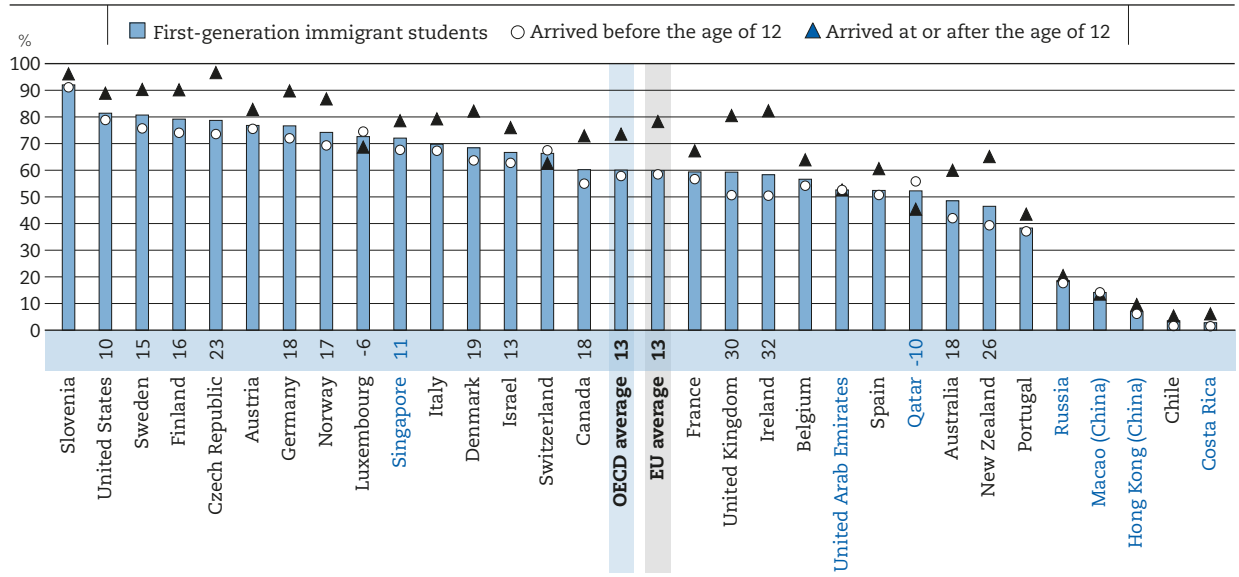
Statistically significant differences in the percentage of non-native-speakers among immigrant students and immigrant students with at least one native-born parent are reported nex to country/economy names.

Countries and economies are ranked in descending order of the percentage of immigrants students who do not speak the language of assessment at home.

Source: OECD, PISA 2015 Database, Table 5.4.

StatLink http://dx.doi.org/10.1787/888933681502

Figure 5.3c ▪ **Percentage of non-native speakers, by age at arrival**



Notes: Non-native-speakers are students who reported that the language they most frequently speak at home is different from the language of the PISA assessment.

Only countries with valid values for first-generation immigrants who arrived before the age of 12 and those who arrived at or after the age of 12 are shown.

Statistically significant differences in the percentage of non-native-speakers among students who arived at or after the age of 12 and among those who arrived before the age of 12 are reported next to country/economy names.

Countries and economies are ranked in descending order of the percentage of first-generation immigrant students who do not speak the language of assessment at home.

Source: OECD, PISA 2015 Database, Table 5.6.

StatLink http://dx.doi.org/10.1787/888933681521

Language barriers and the academic performance of students with an immigrant background

PISA reveals that speaking a language at home that is different from the language of assessment can explain much of the difference in academic performance between native students and students with an immigrant background. First and foremost, it affects students' proficiency in reading, which in turn also influences their competence in other domains. One way to examine the association between the language spoken at home and students' performance is to analyse students' performance in reading relative to their performance in mathematics. By controlling for mathematics performance, students' language-specific abilities are isolated from their general information-processing skills.

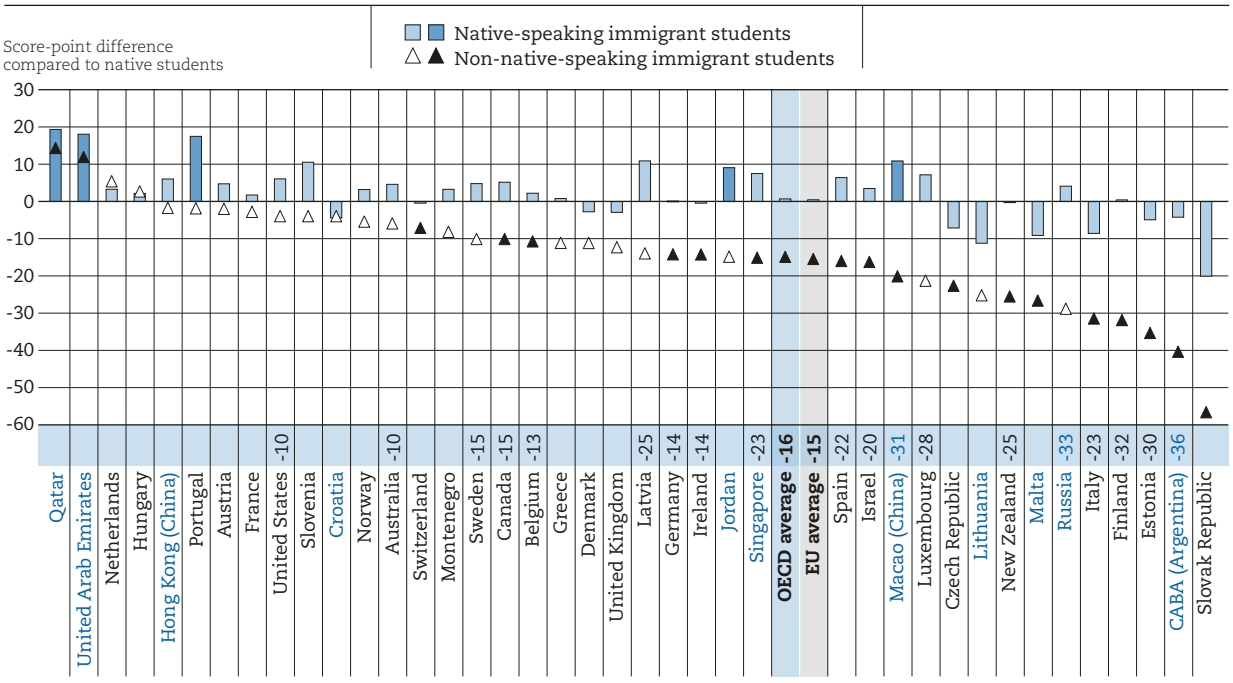
Figure 5.4 compares the reading scores of native students who speak the language of assessment at home and of immigrant students who do, and do not, speak the language of assessment at home, after accounting for their mathematics scores (results for other groups of students with an immigrant background can be found in Table 5.7). Significant differences between immigrant students who do and those who do not speak the language of assessment at home are reported next to country names. In almost all countries and economies, in 2015 there was no significant difference in reading scores between immigrant and native students who reported speaking the language of assessment at home after accounting for their mathematics performance and their socio-economic status. In Macao (China), Jordan, Portugal, Qatar and the United Arab Emirates, immigrant students who reported speaking the language of assessment at home scored higher than native students who reported the same, after accounting for the two variables.

By contrast, in most countries, immigrant students who reported that they did not speak the language of assessment at home had lower scores in reading than both native students and immigrant students who reported speaking the language of assessment at home, after accounting for their mathematics scores and socio-economic status. On average across OECD countries, the gap in reading scores between native-speaking and non-native-speaking immigrant students was 16 points (16 points across EU countries);

but differences can be much starker. For instance, in the Slovak Republic, the score-point difference in reading performance between immigrant students who reported that they did not speak the language of assessment at home and native students who reported that they did was 56 points. In Ciudad Autónoma de Buenos Aires (Argentina) (hereafter "CABA [Argentina]"), Estonia, Finland and the Russian Federation (hereafter "Russia"), non-native-speaking immigrant students also scored much lower in reading than both native students and immigrant students who reported speaking the language of assessment at home.

Figure 5.4 ▪ **Difference in reading scores, by immigrant background and language spoken at home**

After accounting for performance in mathematics and socio-economic status



Notes: Statistically significant differences are marked in a darker tone.

Native students are students without an immigrant background who speak most frequently at home the language of the PISA assessment.

Native-speaking students are students who speak most frequently at home the language of the PISA assessment.

Non-native-speaking students are those who reported that the language they most frequently speak at home is different from the language of the PISA assessment.

Only countries with valid values for both native- and non-native-speaking immigrant students are shown.

Statistically significant differences between non-native- and native-speaking immigrant students are shown next to country/economy names. For the OECD and EU averages, this number refers only to the subset of countries/economies with valid information on both groups of students.

Countries and economies are ranked in descending order of the gap in reading scores between non-native-speaking immigrant students and native students, adjusted for socio-economic status and performance in mathematics.

Source: OECD, PISA 2015 Database, Table 5.7.

StatLink http://dx.doi.org/10.1787/888933681540

Language fluency is associated with students' proficiency in all academic domains, even those with less language content. Table 5.8 (available on line) shows score-point differences between native students and different groups of students with an immigrant background in mathematics performance, after accounting for socio-economic status. PISA results show, for example, that in the majority of countries and economies considered, immigrant students who reported that they did not speak the language of assessment at home scored lower in mathematics than native-speaking immigrant students. While the OECD average gap between the two groups was 9 score points, in Croatia, Germany, Hong Kong (China), Jordan, Latvia, Luxembourg, Macao (China), Russia and Switzerland, the gap was more than 25 points wide. In Croatia, Hong Kong (China), Ireland and Russia, immigrant students who reported speaking the language of assessment at home were not at a disadvantage compared to native students. However, immigrant students who reported that they did not speak the language of assessment at home scored much lower than native students.

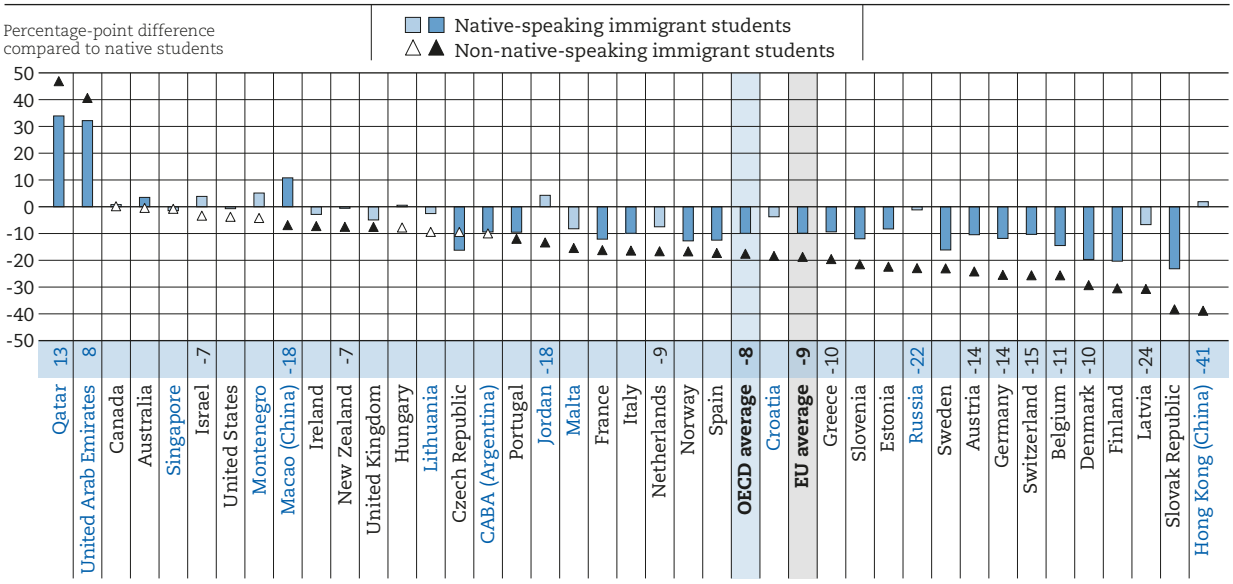
In the majority of countries, non-native-speaking students of mixed-heritage and returning foreign-born students perform as well as native students in reading (Table 5.7, available on line). But, as shown in Table 5.8 (available on line), mathematics scores tell a different story. In the majority of countries and economies, non-native-speaking students of mixed heritage and returning foreign-born students (who together constitute the group of students identified as immigrant students who have at least one native-born parent) scored lower in mathematics than native-speaking immigrant students. On average across OECD countries, the performance gap in mathematics was 28 score points (26 points across EU countries) – about three times larger than the performance gap between native students and immigrant students who have two foreign-born parents. This gap was greater than 25 score points in 20 out of 21 countries where the language penalty is statistically significant. These results indicate that the impact of difficulties with the host-country language on returning foreign-born students and native students of mixed-heritage on mathematics scores seems to be greater than for immigrant students with two foreign-born parents.

Attaining baseline levels of proficiency in the core PISA subjects

PISA shows that whether or not immigrant students speak the language of the assessment at home is strongly associated with their overall academic performance. Indeed, the language spoken at home is significantly associated with students' likelihood of reaching the baseline levels of proficiency in the three core PISA subjects: reading, mathematics and science. Figure 5.5a compares the percentage of immigrant students who are academically resilient among immigrants who do and those who do not speak the language of assessment at home with the percentage of native students who attain that level of proficiency.

Figure 5.5a ▪ **Students attaining baseline academic proficiency, by immigrant background and language spoken at home**

After accounting for socio-economic status



Notes: Statistically significant differences are marked in a darker tone.

Students who attain baseline academic proficiency are those who attain at least proficiency Level 2 in all three core PISA subjects: science, reading and mathematics.

Native students are students without an immigrant background who speak most frequently at home the language of the PISA assessment.

Native-speaking students are students who speak most frequently at home the language of the PISA assessment.

Non-native-speaking students are those who reported that the language they most frequently speak at home is different from the language of the PISA assessment.

Only countries with valid values for both native- and non-native-speaking immigrant students are shown.

Statistically significant differences between non-native- and native-speaking immigrant students with at least one native-born parent are shown next to country/economy names. For the OECD and EU averages, this number refers only to the subset of countries/economies with valid information on both groups of students.

Countries and economies are ranked in descending order of the gap between non-native-speaking immigrant students and native students in the percentage of students attaining baseline academic proficiency, adjusted for socio-economic status.

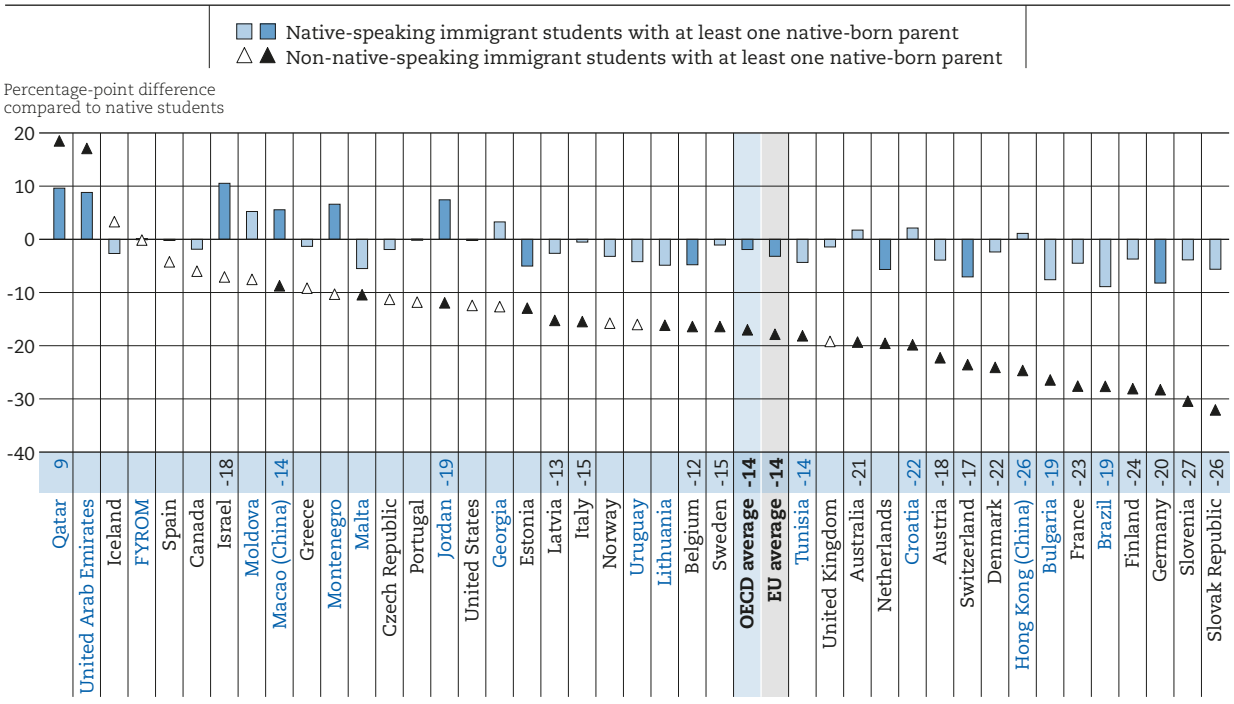
Source: OECD, PISA 2015 Database, Table 5.9.

StatLink http://dx.doi.org/10.1787/888933681559

In most countries and economies, the percentage of non-native-speaking immigrant students who perform at proficiency Level 2 or higher in reading, mathematics and science is lower than the percentage of native students who do. On average in 2015 across OECD countries, 18 percentage points separated the two groups (19 percentage points across EU countries); but in Finland, Hong Kong (China), Latvia and the Slovak Republic, the difference was 30 percentage points or more. In 14 countries and economies, immigrant students who were non-native speakers were less likely to attain baseline levels of proficiency compared to native-speaking immigrant students. The differences between the two groups in the percentage of students who attained the baseline level of proficiency in the three core PISA subjects ranged from 7 percentage points to 41 percentage points. In Croatia, Hong Kong (China), Ireland, Jordan, Latvia, Malta, the Netherlands, New Zealand and Russia and the United Kingdom native students and immigrant students who are native-language speakers stood an equal chance of reaching baseline levels of proficiency in all three subjects, while immigrant students who are non-native speakers were significantly less likely to achieve the same result. In these countries, fluency in the language of assessment is key to whether immigrant students attain the baseline level of proficiency in reading, mathematics and science, after accounting for socio-economic status.

Figure 5.5b ▪ **Students attaining baseline academic proficiency, by immigrant heritage and language spoken at home**

After accounting for socio-economic status



Notes: Statistically significant differences are marked in a darker tone.

Students who attain baseline academic proficiency are those who attain at least proficiency Level 2 in all three core PISA subjects: science, reading and mathematics.

Native students are students without an immigrant background who speak most frequently at home the language of the PISA assessment.

Native-speaking students are students who speak most frequently at home the language of the PISA assessment.

Non-native-speaking students are those who reported that the language they most frequently speak at home is different from the language of the PISA assessment.

Only countries with valid values for both native- and non-native-speaking immigrant students with at least one native-born parent are shown.

Statistically significant differences between non-native- and native-speaking immigrant students with at least one native-born parent are shown next to country/economy names. For the OECD and EU averages, this number refers only to the subset of countries/economies with valid information on both groups of students.

Countries and economies are ranked in descending order of the gap between non-native-speaking immigrant students with at least one native-born parent and native students in the percentage of students attaining baseline academic proficiency, adjusted for socio-economic status.

Source: OECD, PISA 2015 Database, Table 5.9.

StatLink http://dx.doi.org/10.1787/888933681578

Figure 5.5b shows that the impact of language fluency is similar, if not stronger, among immigrant students who have at least one native-born parent. On average across OECD countries, the difference in the percentage of immigrant students in this group reaching baseline levels of proficiency between students who do and those who do not speak the language of assessment at home is 14 percentage points (14 across EU countries) – a larger difference than among immigrant students with two foreign-born parents. In 14 countries and economies, native-speaking immigrant students with at least one foreign-born parent are as likely as native students to attain the baseline level of proficiency in PISA subjects; but non-native-speaking students within this group are not. In Bulgaria, Finland, France, the Slovak Republic and Slovenia, the difference compared to native students is more than 25 percentage points for non-native speakers, and it is not statistically significant for native speakers. These results suggest that fluency in the language of assessment is linked to academic resilience among immigrant students with at least one foreign-born parent.

Linguistic barriers to academic achievement vary significantly across the four previously defined categories of students with an immigrant background (first-generation immigrant students, second-generation immigrant students, returning foreign-born students, and native students of mixed heritage) (Table 5.10, available on line). On average across OECD countries and in most countries and economies, the students most affected by being non-native speakers as opposed to native speakers are returning foreign-born students, followed by first-generation immigrant students, native students with a mixed heritage and finally second-generation immigrant students. However, as shown in Figure 5.6, there are significant differences across countries. In France, for example, non-native-speaking returning foreign-born students are 33% less likely than native students to reach baseline levels of proficiency in reading, mathematics and science (a 26 percentage-point difference compared to native-speaking returning foreign-born students), while other groups of students with an immigrant background show less or non-significant language-related impact on their performance.

In Denmark, proficiency in the host-country language plays a much larger role among second-generation immigrant students and native students of mixed heritage. Among the former category of students, the language-related difference in the likelihood of students attaining baseline levels of academic achievement is 11 percentage points in favour of those who speak the language of assessment. Among the latter group, the difference is 23 percentage points. There is no statistically significant difference between native- and non-native speakers in the likelihood of first-generation immigrant students and returning foreign-born students attaining baseline levels of academic proficiency.

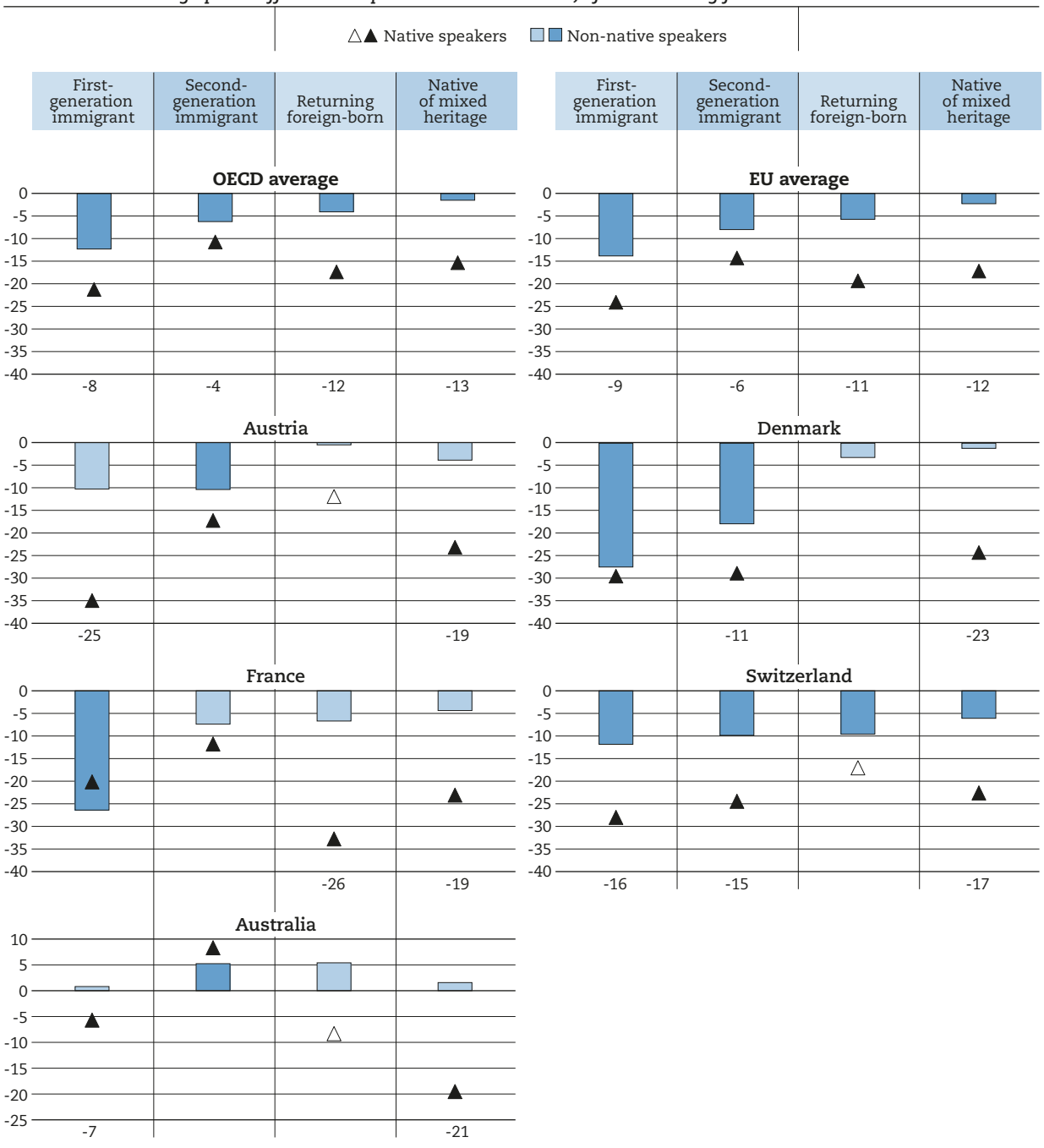
In Austria, the effect of speaking a language at home that is different from the one of assessment is large for first-generation immigrant students and native students of mixed heritage, while it is not significant for other groups of students with an immigrant background. In Switzerland, the language-related disadvantage is similar for all groups of students with an immigrant background except for returning foreign-born students, but that is probably due to small sample-size effects.

In Australia, native-speaking first-generation immigrant students and native students of mixed heritage have the same probability of attaining baseline proficiency in the three core PISA subjects as native students. However, non-native-speaking students among these groups of students with an immigrant background show significant disadvantages compared to native students and native-speaking students. Caution is advised when analysing results in Figure 5.6, because the sample of the groups considered might be small and varies across the different groups.

Language barriers also explain part of the relatively low performance among first-generation immigrant students related to their age at arrival in the host country. Table 5.6 (available on line) shows that in the majority of countries and economies, late arrivals are more likely than early arrivals to speak at home a language that is different from the language of assessment. Figure 5.7 shows that in several countries and economies, fluency in the host-country language accounts for a significant portion of the performance gaps between early arrivals (students who had arrived before the age of 12) and late arrivals (students who had arrived at or after the age of 12). On average across OECD countries, the difference between the two groups of students in the percentage of students who attain baseline proficiency in all three core PISA subjects is reduced by 5 percentage points (from 16 percentage points to 11 percentage points) after the language spoken at home is taken into account. Across EU countries, the difference shrinks by almost 5 percentage points (from around 17 to 12 percentage points).

Figure 5.6 ▪ **Students attaining baseline academic proficiency, selected countries**

Percentage-point difference compared to native students, after accounting for socio-economic status



Notes: Statistically significant differences are marked in a darker tone.

Students who attain baseline academic proficiency are those who attain at least proficiency Level 2 in all three core PISA subjects: science, reading and mathematics.

Native students are students without an immigrant background who speak most frequently at home the language of the PISA assessment.

Native-speaking students are students who speak most frequently at home the language of the PISA assessment.

Non-native-speaking students are those who reported that the language they most frequently speak at home is different from the language of the PISA assessment.

Statistically significant differences between non-native- and native-speaking students with an immigrant background are shown at the bottom of the panels. For the OECD and EU averages, this number refers only to the subset of countries/economies with valid information on both groups of students.

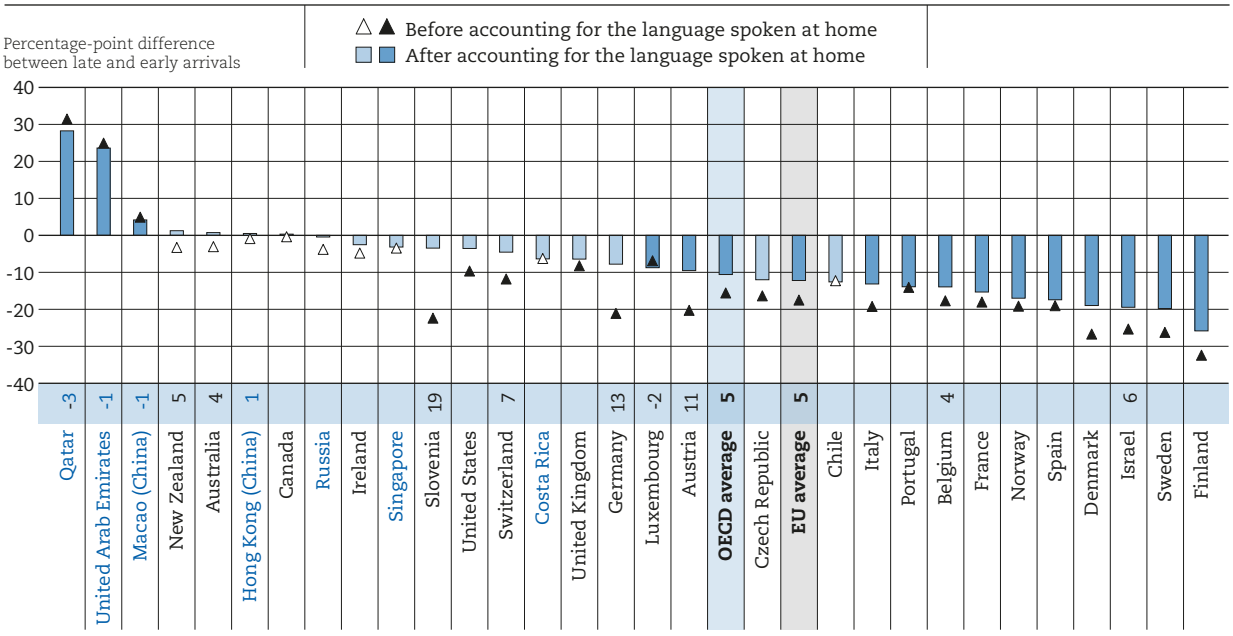
Source: OECD, PISA 2015 Database, Table 5.10.

StatLink http://dx.doi.org/10.1787/888933681597

In Austria, Germany and Slovenia, these differences are reduced by more than 10 percentage points. In Slovenia, linguistic differences explain a large part of the late arrival penalty: after the effect of speaking a different language is accounted for, the gap between late and early arrivals is reduced by 19 percentage points (from 22 percentage points to a not statistically significant 3 percentage points).

Figure 5.7 ▪ **Difference between late and early arrivals in attaining baseline academic proficiency**

After accounting for socio-economic status



Notes: Statistically significant differences are marked in a darker tone.

Early arrivals are first-generation immigrant students who arrived at the country where they sat the PISA test before the age of 12. Late arrivals are first-generation immigrant students who arrived at the country where they sat the PISA test at or after the age of 12.

Statistically significant differences in the late arrival gap before and after accounting for socio-economic status are shown next to the country/economy names.

Countries and economies are ranked in descending order of the percentage-point difference between late arrivals and early arrivals in the percentage of students attaining baseline academic proficiency, after accounting for the language spoken at home and socio-economic status.

Source: OECD, PISA 2015 Database, Table 5.11.

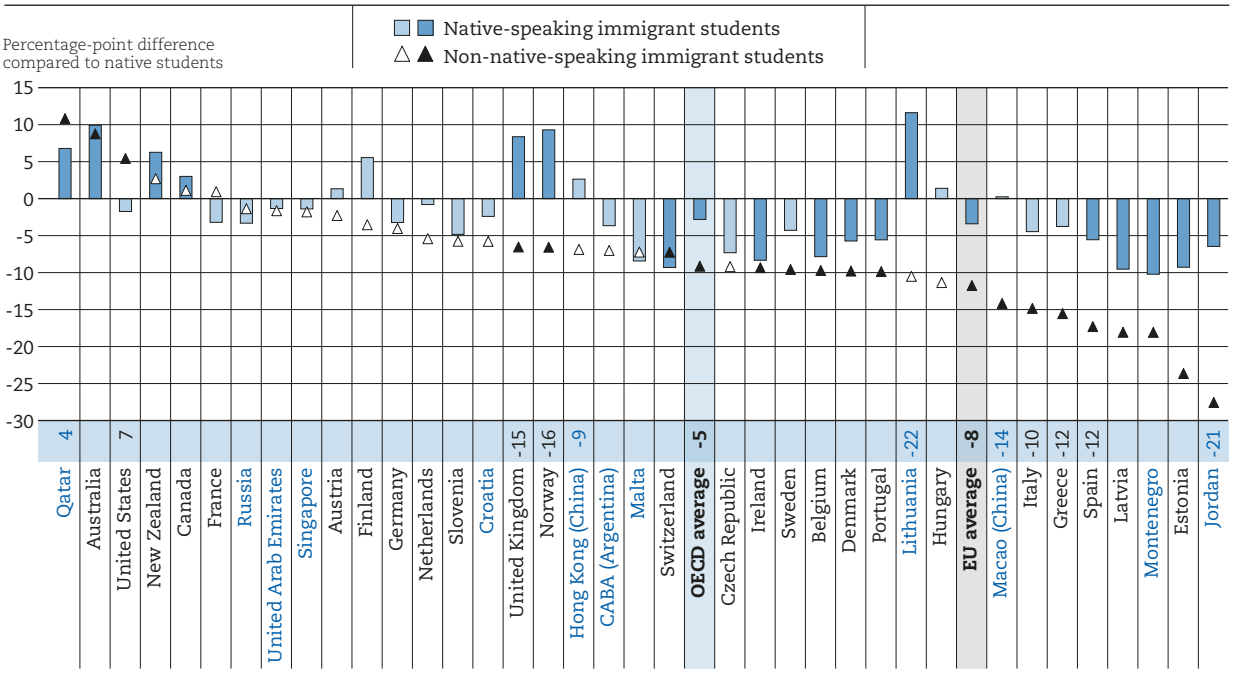
StatLink http://dx.doi.org/10.1787/888933681616

Language barriers and the well-being of students with an immigrant background

Speaking a language at home that is different from the host-country language has a significant impact on the sense of belonging reported by students with an immigrant background. Figure 5.8a shows differences in the percentage of immigrant students who attain a baseline sense of belonging – i.e. who feel that they belong at school[3] – across the groups defined in previous sections (non-native-speaking immigrant students, native-speaking immigrant students and native students) (results for all groups of students with an immigrant background can be found in Table 5.13 available on line). Figure 5.8a shows that in 2015, in a large number of countries and economies, non-native-speaking immigrant students were less likely than native and native-speaking immigrant students to feel like they belong at school. They were also less likely to be socially resilient. On average across OECD countries, the share of students who reported a sense of belonging was five percentage points smaller among non-native-speaking immigrant students than among native-speaking immigrant students, and nine percentage points smaller than among native students (six and twelve percentage points, respectively, across EU countries). In several countries and economies, native-speaking immigrant students had equal or higher chances of reporting a sense of belonging at school compared to native students, and had significantly greater chances compared to non-native-speaking immigrant students. This is the case in Greece, Italy, Macao (China) and Sweden, while in Norway and the United Kingdom, native-speaking immigrant students were more likely to report that they feel like they belong at school, even compared to native students.

Figure 5.8a ■ **Students reporting a sense of belonging at school, by immigrant background and language spoken at home**

After accounting for socio-economic status



Notes: Statistically significant differences are marked in a darker tone.

Students who reported a sense of belonging at school are those who reported that they "agree" or "strongly agree" with the statement "I feel like I belong at school" and "disagree" or "strongly disagree" with the statement "I feel like an outsider at school".

Native students are students without an immigrant background who speak most frequently at home the language of the PISA assessment.

Native-speaking students are students who speak most frequently at home the language of the PISA assessment.

Non-native-speaking students are those who reported that the language they most frequently speak at home is different from the language of the PISA assessment.

Only countries with valid values for both native- and non-native-speaking immigrant students are shown.

Statistically significant differences between non-native- and native-speaking immigrant students with at least one native-born parent are shown next to country/economy names. For the OECD and EU averages, this number refers only to the subset of countries/economies with valid information on both groups of students.

Countries and economies are ranked in descending order of the gap between non-native-speaking immigrant students and native students in the percentage of students who reported a sense of belonging at school, adjusted for socio-economic status.

Source: OECD, PISA 2015 Database, Table 5.12.

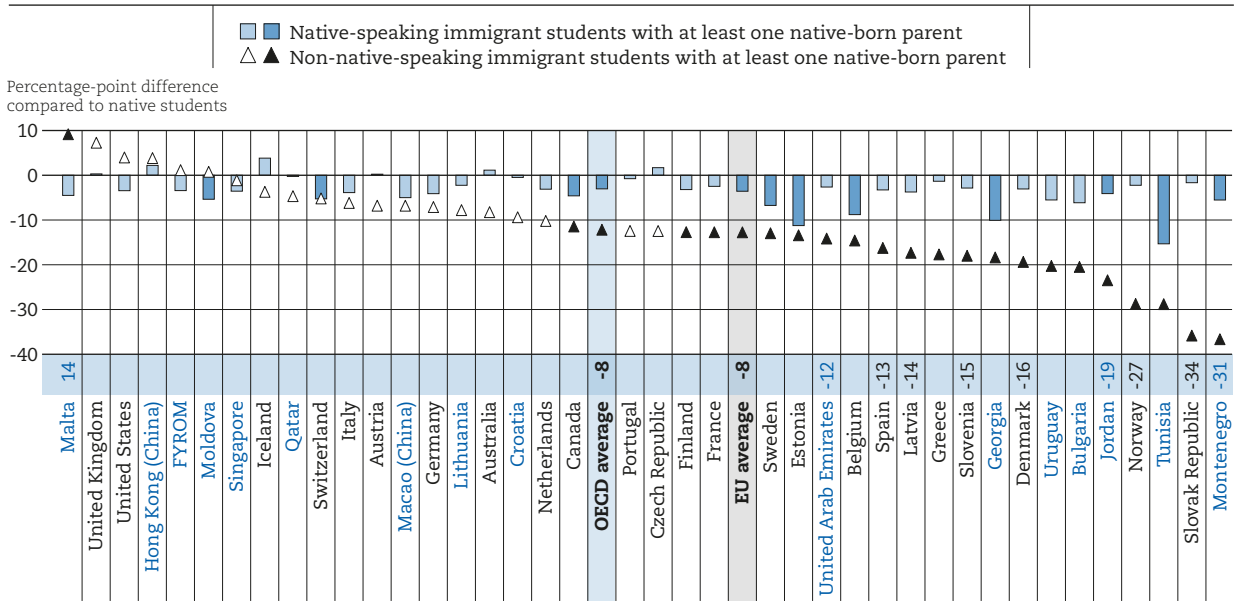
StatLink http://dx.doi.org/10.1787/888933681635

Figure 5.8b reports the same differences in sense of belonging but for immigrant students with at least one native-born parent (a category which includes returning foreign-born students and native students of mixed heritage). On average across OECD, the share of students who reported that they feel like they belong at school among non-native-speaking students was eight percentage points smaller than the share among native-speaking students (eight percentage points across EU countries), and 12 percentage points smaller than the share among native students (13% across EU countries). In Bulgaria, Latvia, Norway, Slovenia, Spain, the United Arab Emirates and Uruguay, native-speaking immigrant students with at least one native-born parent were as likely to report a sense of belonging at school as native students, but much more likely than non-native-speaking immigrant students with at least one native-born parent. In Montenegro, Norway and the Slovak Republic, the difference between native-speaking and non-native-speaking immigrant students with at least one native-born parent in the shares of students who reported that they feel like they belong at school was larger than 25 percentage points. Overall, language seems to play a more decisive role in determining students' sense of belonging among immigrant students with at least one native-born parent than among immigrant students with two foreign-born parents.

Table 5.13 (available on line) examines language-related differences in the sense of belonging at school for each of the four categories of students with an immigrant background. Several effects are not statistically significant because of issues related to sample size. Nonetheless, some interesting results emerge.

Figure 5.8b ▪ **Students reporting a sense of belonging at school, by immigrant heritage and language spoken at home**

After accounting for socio-economic status



Notes: Statistically significant differences are marked in a darker tone.

Students who reported a sense of belonging at school are those who reported that they "agree" or "strongly agree" with the statement "I feel like I belong at school" and "disagree" or "strongly disagree" with the statement "I feel like an outsider at school".

Native students are students without an immigrant background who speak most frequently at home the language of the PISA assessment.

Native-speaking students are students who speak most frequently at home the language of the PISA assessment.

Non-native-speaking students are those who reported that the language that they most frequently speak at home is different from the language of the PISA assessment.

Only countries with valid values for both native- and non-native-speaking immigrant students with at least one native-born parent are shown.

Statistically significant differences between non-native- and native-speaking immigrant students with at least one native-born parent are shown next to country/economy names. For the OECD and EU averages, this number refers only to the subset of countries/economies with valid information on both groups of students.

Countries and economies are ranked in descending order of the gap between non-native-speaking immigrant students with at least one native-born parent and native students in the percentage of students who reported a sense of belonging at school, adjusted for socio-economic status.

Source: OECD, PISA 2015 Database, Table 5.12.

StatLink http://dx.doi.org/10.1787/888933681654

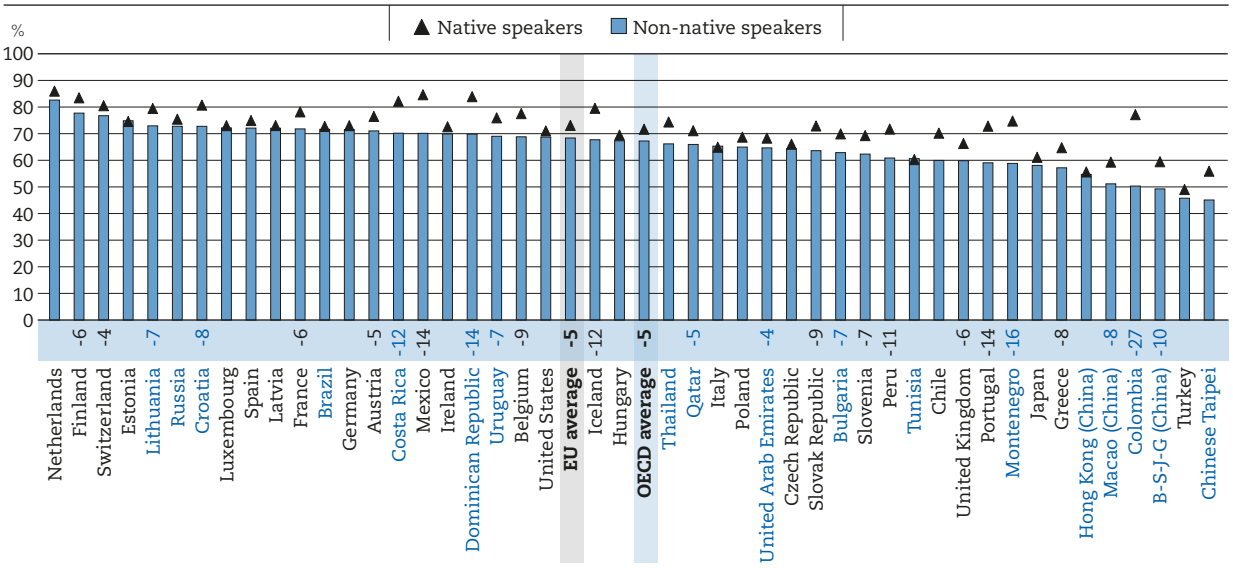
Linguistic differences were particularly important in Greece, Macao (China), Norway, Sweden and the United Kingdom, where native-speaking first-generation immigrant students were as likely to report a sense of belonging as native students, but non-native speakers within this group were significantly less likely to report so. In all five countries, the difference in the percentages of native- and non-native speaking first-generation immigrant students who reported that they feel like they belong at school was more than 10 percentage points. Curiously, in Norway and the United Kingdom, native-speaking second-generation immigrant students were more likely than native students to feel like they belong at school, but the same was not true among their non-native-speaking peers. By contrast, in the United Kingdom, non-native-speaking native students of mixed heritage were more likely to report a sense of belonging compared to native students and native-speaking students with the same immigrant background.

Speaking a language at home that is different from the language of instruction is also associated with the emotional well-being of students, as measured by their satisfaction with life and self-reported levels of anxiety. Figure 5.9 shows the percentage of students who reported being satisfied with life[4] among students who do and those who do not speak the language of assessment at home. In the great majority of countries, students who speak the language of assessment at home were more likely to report being satisfied with life than students who speak a different language (detailed results for each group of students with an immigrant background can be found Table 5.14 available on line). On average across OECD countries, the share of students who reported being satisfied with life was 5 percentage points

larger among students who speak the language of instruction at home than the share among students who do not speak the language of instruction at home. In B-S-J-G (China), Colombia, Costa Rica, Dominican Republic, Iceland, Mexico, Montenegro, Peru and Portugal, the difference between the two groups was greater than 10 percentage points.

Figure 5.9 ▪ **Life satisfaction, by language spoken at home**

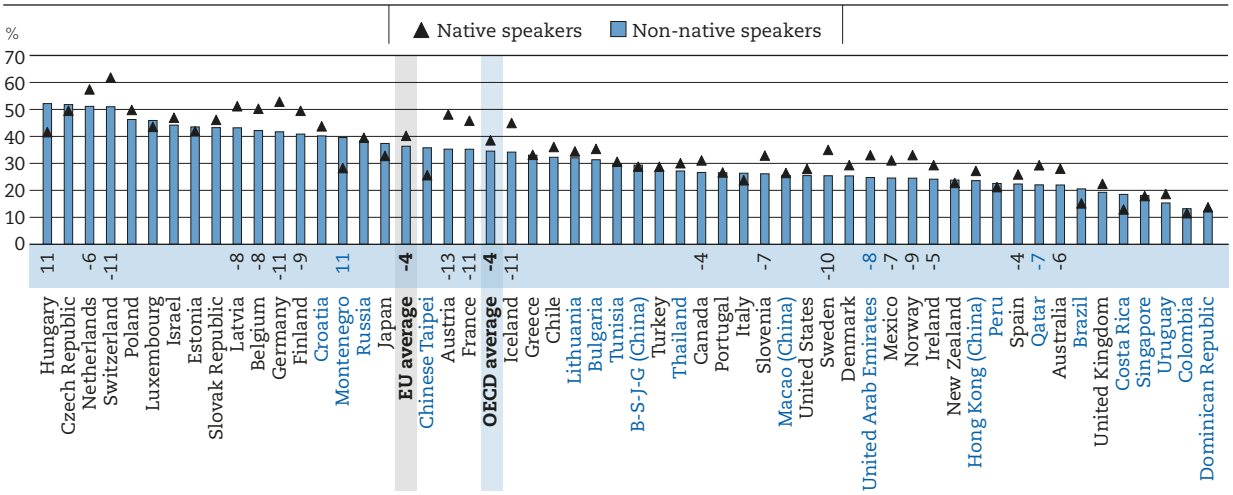
Percentage of students who reported being satisfied with life



Notes: Statistically significant differences between non-native and native-speakers are shown next to the country/economy names.
Only countries/economies with valid data on native and non-native speakers are shown.
Students who reported being satisfied with life are those who reported a life satisfaction of 7 or above on a scale from 0 to 10.
Countries and economies are ranked in descending order of the percentage of non-native-speaking students who reported that they are satisfied with life.
Source: OECD, PISA 2015 Database, Table 5.14.
StatLink http://dx.doi.org/10.1787/888933681673

Figure 5.10 ▪ **Low schoolwork-related anxiety, by language spoken at home**

Percentage of students who reported low levels of schoolwork-related anxiety



Notes: Statistically significant differences between non-native and native-speakers are shown next to the country/economy names.
Only countries/economies with valid data on native and non-native speakers are shown.
Students who reported low schoolwork-related anxiety are those who reported that they "disagree" or "strongly disagree" with the statements "I often worry that it will be difficult for me taking a test" and "Even if I am well prepared for a test, I feel very anxious".
Countries and economies are ranked in descending order of the percentage of non-native-speaking students who reported low levels of schoolwork-related anxiety.
Source: OECD, PISA 2015 Database, Table 5.14.
StatLink http://dx.doi.org/10.1787/888933684409

Figure 5.10 shows the percentage of students who reported low schoolwork-related anxiety[5] among students who speak the language of instruction at home and those who do not (detailed results for each group of students with an immigrant background can be found Table 5.14 available on line). In several countries, speaking a different language at home was associated with a greater chance of being anxious. On average across OECD and EU countries, the difference between the percentage of students who reported speaking and those who reported not speaking the language of assessment at home and who reported low schoolwork-related anxiety was four percentage points. In Austria, France, Germany, Iceland, Sweden and Switzerland, the difference between the two groups was larger than 10 percentage points.

Language groups and linguistic distance

The figures and analyses presented so far report information on the linguistic background of students based on whether a student reported speaking the language of the assessment at home or not in 2015. But this information says nothing about the composition of the non-native language group, which can be diverse. Figure 5.11 shows some of the different linguistic groups that are included in the population of students with an immigrant background in a selected group of PISA countries. Specifically, it focuses on students in this group who reported speaking a language at home that is different from the one in which they were assessed. The countries shown vary considerably in the languages spoken at home by students with an immigrant background.

In Austria and Denmark, there were more than ten well-defined language groups. In both countries, 20% of students with an immigrant background were included in one language group; the rest were distributed across other groups. In Luxembourg, there were fewer language groups; but one of them – Portuguese – was spoken by 39% of students with an immigrant background who reported speaking a language at home that is different from the language in which they sat the PISA assessment. In another set of countries, most students with an immigrant background belonged to a single language group. This is the case in Estonia and the United States, where Russian and Spanish, respectively, were the most commonly spoken languages among students with an immigrant background who reported that they do not speak the language of assessment at home.

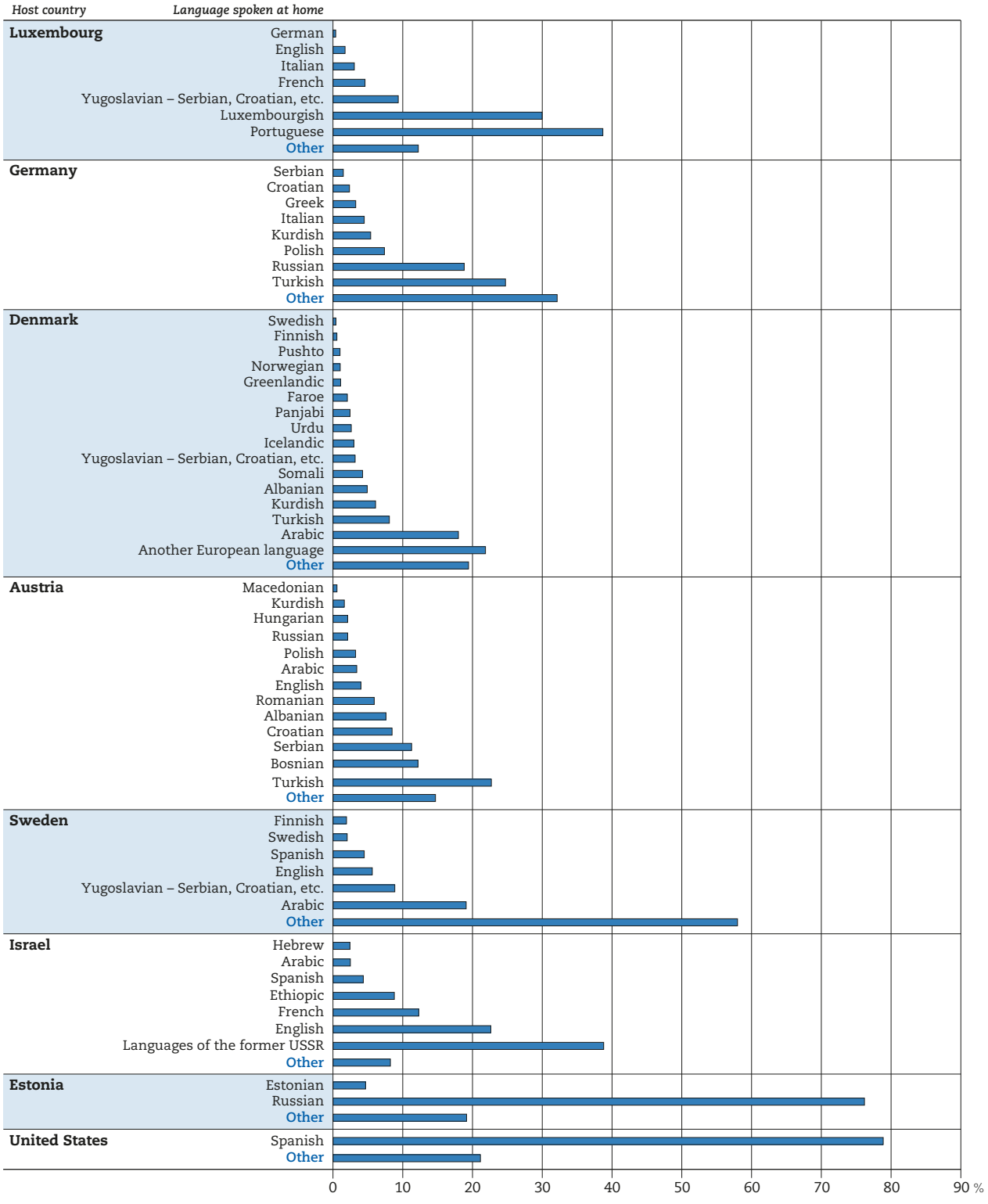
Given the wide spectrum of languages that students with an immigrant background speak in PISA-participating countries, it is insufficient, for research and policy purposes, to distinguish solely between students who predominantly speak, at home, the same language as that of the assessment and those who speak a different language. Languages can be more or less similar to one another; therefore, identifying the specific language that an immigrant speaks can shed light on the magnitude of the language penalty he or she faces. For example, the language barrier that immigrants from Spanish-speaking countries face when settling in Italy is not the same as the one that Italian-speaking immigrants face when they settle in Finland.

This section quantifies the association between students' outcomes and the degree of linguistic proximity between the language that students primarily speak at home and the language in which they sat the PISA test. It builds on the observation made in previous sections of this chapter that students with an immigrant background who report speaking the language of assessment at home are more likely to be academically, socially and emotionally resilient than those who do not. The following analysis aims to establish if the relative difficulty in learning a distal language explains differences in outcomes among students with an immigrant background, especially among immigrant students who arrived after the age of 12, a critical age for acquiring language proficiency.

To investigate such relationships, this section uses a lexicostatistical measure of linguistic proximity. The Automated Similarity Judgment Program (ASJP) was developed by the German Max Planck Institute for Evolutionary Anthropology. It is based on a comparison of the pronunciation of words that have the same meaning in pairs of languages, using a composite Levensthein distance indicator. Box 3.1 explains how this measure is computed, as detailed in Bakker et al. (2009). In order to determine whether the variability in the outcomes of students with an immigrant background is associated with how similar/dissimilar their mother tongue is to the language of their host country, the linguistic proximity indicator was computed for all pairs of languages that appear in the PISA student sample (see Table 5.15 for a sample of language pairs). The Language Distance Index (LDI) is then used as a control in models that estimate differences in academic and general well-being outcomes.

Figure 5.11 ▪ **Non-native-speaking students with an immigrant background, by language spoken at home**

Selected OECD countries



Note: Non-native-speaking students are students who reported that the language they most frequently speak at home is different from the language of the PISA assessment.

Source: OECD, PISA 2015 Database.

StatLink http://dx.doi.org/10.1787/888933684428

Languages can differ in a number of aspects: vocabulary, grammar, pronunciation, scripture and phonetic inventories. The overall distance between any two languages reflects the degree of dissimilarity across all key aspects and the ease/difficulty with which individuals speaking one language can acquire proficiency and mastery in the other language. A key limitation of the ASJP is that the resulting indicator only captures differences in pronunciation between languages. This is particularly relevant because the ASJP measures the distance between languages only in their spoken form, while the PISA measures are based on a written text. However, the main advantage of the ASJP measure is that it is easily computable for any pair of languages and thus readily available for analyses. For this reason, it is the most widely used measure of linguistic distance in empirical work (Bakker et al., 2009; Isphording and Otten, 2014).

Box 5.1. **Estimation of the language distance index: Levensthein distance and the Automated Similarity Judgment Program**

The Levensthein distance is a metric developed to identify the difference between two sequences. When comparing words, the Levenshtein distance characterises the minimum number of single-character edits (insertions, deletions or substitutions) that one is required to perform in order to change one word into the other (Levenshtein, 1966). This section is based on previous work exploiting the Levenshtein distance to compute the level of dissimilarity across combinations of languages (Bakker et al., 2009).

The Max Planck's ASJP developed a composite indicator based on the automatic comparison of the pronunciation of 40 words that have the same meaning from 4 664 languages. The indicator is built using the following procedure. First, each pair of words i with the same meaning is judged according to the similarity in pronunciation, by counting the number of insertions, deletions or substitutions of consonants and vowels that are necessary to transfer the phonetic transcription of one word (in language x) to the other correspondent word in language y, obtaining a measure of the distance between language x and y for the pair of words i, D_i (x, y). For example, the English word person – expressed phonetically as *pers3n* – needs two insertions, deletions or substitutions to be transformed into the same word in Spanish, *persona*.

To aid interpretations, the table below displays some examples of the Levenshtein distance between words of different languages with the same meaning. This first value estimated, for each pair of words, is then normalised by the potential maximum distance between both words, obtaining D_i^n (x, y). The average the normalised distances estimated for the 40 words in the list is then computed obtaining the normalised language distance between languages x and y, $LD_{(x,y)} = \Sigma_i D_i (x, y) \frac{1}{M}$. This estimate is normalised again by dividing it by the global distance $T_{(x,y)} = \Sigma_i \neq_j D(x_i, y_j) \frac{1}{M(M-1)}$, which is the average distance between any word in the list in language α with any word in the list for language β.

Finally, to obtain a definitive measure of language distance, the previous vale of normalised language distance $LD_{(x,y)}$ is divided by the global distance $T_{(x,y)}$ to obtain the normalised and divided Levensthein distance $LDND_{(x,y)} = \frac{LD_{(x,y)}}{T_{(x,y)}}$.

For example, the Levensthein distance between Spanish and English for the word "you" equals 1 because in Spanish the word you is pronounced as "tu" while in English it is pronounced as "yu" (requiring one substitution). The Levensthein distance between Spanish and English for the word "night" equals 3 because in Spanish the word night is pronounced as "noCe" while in English it is pronounced as "nEit" (requiring three substitutions).

Source: Brown et al. (2008).

Table 5.15 shows large variations in linguistic proximity for non-native speakers, ranging from 19.60 – the distance between Serbian and Croatian – to 104.06, the estimated dissimilarity between English and Vietnamese. Within countries, there are significant differences in the language distance that various groups of immigrants face. For example, among immigrants who live in Spain, the language distance faced by students who speak Catalan at home and who sat the PISA test in Valencian is considerably different from the one faced by students who speak Basque at home and who sat the test in Valencian. The language-distance index for the first pair of languages is 26.34, while it is 103.15 for the second pair.

Table 5.15 ▪ **Linguistic distance between the PISA test language and languages spoken at home**

	Maximum LDI			Minimum LDI			
	Test language	Language spoken at home	Value	Test language	Language spoken at home	Value	Difference
Spain	Valencian	Basque	103.15	Valencian	Catalan	26.34	76.81
Croatia	Croatian	Hungarian	94.37	Croatian	Serbian	19.60	74.77
Czech Republic	Czech	Chinese	102.01	Czech	Slovak	32.82	69.19
FYROM	Turkish	Serbian	97.97	Macedonian	Serbian	36.59	61.38
Finland	Swedish	Thai	100.99	Finnish	Estonian	41.93	59.06
Luxembourg	French	German	97.01	German	Luxembourgish	41.94	55.07
Denmark	Danish	Somali	102.27	Danish	Norwegian	47.85	54.42
Belgium	Dutch	Turkish	102.33	Dutch	German dialect (BEL)	48.83	53.50
Switzerland	German	Turkish	100.05	Italian	Spanish	58.31	41.74
Latvia	Latvian	Ukrainian	92.27	Russian	Belarusian	59.2	33.07
Sweden	English	Arabic	98.03	Swedish	English	66.26	31.77
Australia	English	Vietnamese	104.06	English	German	72.61	31.45
Lithuania	Russian	Lithuanian	91.33	Polish	Russian	61.06	30.27
Moldova	Russian	Romanian	96.68	Russian	Bulgarian	66.64	30.04
Austria	German	Turkish	100.05	German	English	72.61	27.44
Uruguay	Spanish	English	92.25	Spanish	Portuguese	67.96	24.29
Israel	Arabic	Spanish	98.49	Hebrew	Arabic	77.63	20.86
Costa Rica	Spanish	Mandarin	98.66	Spanish	French	84.03	14.63
Germany	German	Turkish	100.05	German	Italian	86.61	13.44
Dominican Republic	Spanish	Creole	97.06	Spanish	French	84.03	13.03
Peru	Spanish	Aymara	102.43	Spanish	English	92.25	10.18
Colombia	Spanish	English	92.25	Spanish	French	84.03	8.22
Georgia	Russian	Azerbaijani	101.15	Azerbaijani	Georgian	93.83	7.32
United Kingdom	English	Scottish Gaelic	96.72	Welsh	Irish	89.56	7.16
Italy	Slovenian	German	92.96	Italian	German	86.61	6.35
Macao (China)	Chinese	Portuguese	101.76	English	Portuguese	95.45	6.31
Lebanon	English	Arabic	98.03	English	French	92.06	5.97
Slovenia	Slovenian	Hungarian	93.93	Slovenian	Italian	88.61	5.32
Korea	Korean	Japanese	99.61	Korean	Chinese	95.91	3.70
New Zealand	English	Chinese	101.27	English	Samoan	97.76	3.51
Singapore	English	Chinese	101.27	English	Malay	98.51	2.76
Hong Kong (China)	Chinese	English	101.27	English	Cantonese	98.52	2.75
Slovak Republic	Hungarian	Romany	98.03	Slovak	Romany	95.81	2.22
Qatar	English	Arabic	98.03	English	Hindi	97.04	0.99
Norway	Nynorsk	Swedish	48.53	Bokmål	Danish	47.85	0.68
United Arab Emirates	English	Arabic	98.03	English	Arabic	98.03	0.00

Source: OECD, PISA 2015 Database and Wichmann, Søren, Eric W. Holman, and Cecil H. Brown (eds.), 2016. The ASJP Database (version 17).
StatLink http://dx.doi.org/10.1787/888933681730

Linguistic distance and students' outcomes

To investigate the effect of language distance on outcomes of interest a series of models were developed and replicated for each measure of resilience: academic, social, emotional and motivational. Results are presented in Tables 5.A1.1 through 5.A1.5 in Annex 5.A1. All analyses were conducted on the pooled PISA 2015 sample, weighting each country for population size and weighting individual students to derive estimates representative at the population level. All regressions included destination-country fixed effects.

Results reported in Table 5.A1.1 indicate that the greater the dissimilarity between the language a student speaks at home and the one in which the student sat the PISA test, the lower his or her likelihood of attaining baseline academic proficiency. Results presented in Model 1 of Table 5.A1.1 indicate that, on average, the percentage of students who reach baseline levels of academic proficiency is 4.3 percentage points lower among students with an immigrant background who speak the language of assessment at home compared to native students. However, the gap is greater among students with an immigrant background who do not speak the language of assessment at home and it increases with the linguistic

distance between the home and assessment languages. A 10-point increase in the linguistic-distance indicator corresponds to a 1.6 percentage-point reduction in the likelihood of reaching baseline levels of academic performance.

Composition effects explain some of the differences between students with and without an immigrant background as well as the gap associated with language distance. When comparing students of the same gender and with similar socio-economic status, the disadvantage related to linguistic distance shrinks (Model 2a). However, once differences in the Human Development Index (HDI) and GDP between countries of origin and host countries are considered, the coefficient associated with having an immigrant background returns to previous levels and the negative effect of linguistic distance grows (Model 2b). In model 2b, a 10-point increase in linguistic distance reduces the probability of reaching baseline academic levels by 1.1 percentage points.

Models 3 and 4 reveal large variations in the extent to which greater linguistic distance reduces the likelihood that 15-year-old students will attain baseline academic proficiency. They show the interaction between the linguistic distance indicator and the socio-economic status and gender of students. Specifically, the negative effect of linguistic distance on the likelihood of reaching baseline levels of performance in the three core PISA subjects tends to be larger among socio-economically advantaged students and among boys. Since 15-year-old boys tend to struggle more than girls with text comprehension (OECD, 2015a), irrespective of their language proficiency, this result appears to suggest that while girls might be able to overcome language barriers, boys might not, which could prevent them from developing academic resilience. Language dissimilarity might represent a greater barrier to advantaged students than to disadvantaged students because disadvantaged students face so many constraints to developing academic resilience that their likelihood of being academically resilient is already low, even before considering whatever language barriers they might face. The more advantaged the student, the more salient the language difficulties in influencing the likelihood that the student will attain baseline levels of proficiency.

Models 5 and 6 focus on immigrant students and identify whether arriving in the host country earlier reduces the negative consequences associated with having a mother tongue that is very different from the language of instruction. Results confirm the critical-period hypothesis: children who had arrived in their host country before the age of 12 have a greater chance of being academically resilient. Indeed, among students who had arrived at or before the age of 12, the share of academically resilient students is 14 percentage points larger than the share of academically resilient students who had arrived after the age of 12. When including age at arrival in the regression (Model 5), the effect of linguistic distance is still statistically significant and large: a one-point increase in the index of linguistic distance reduces the chance of reaching baseline levels of academic achievement by 1.3 percentage points. When also considering the interaction of age at arrival and the linguistic distance indicator (Model 6), the effect of the linguistic-distance variable and its interaction with age at arrival are not statistically significant at conventional levels (5%). Nevertheless, the coefficients are negative, as expected.

Table 5.A1.2 identifies the relationship between linguistic distance and the likelihood of reporting a sense of belonging. When considering only immigrant background, gender, socio-economic status and score on the linguistic distance indicator for students (Models 2a and 2b), the indicator of linguistic distance is either not statistically significant or it has only small effects. However, when also considering HDI and GDP differentials and the interaction of the linguistic-distance indicator with gender and socio-economic status (Models 2b, 3 and 4), the effect of linguistic distance on sense of belonging at school is statistically significant and similar to that on academic resilience. In Model 3, students with an immigrant background who speak the language of assessment at home have a 3.4 percentage points lower chance of reporting a strong sense of belonging than their native peers. However, immigrant students who do not speak the language of assessment at home have even lower likelihoods, which decrease by 1 percentage points with each 10-point increase in linguistic distance.

Unlike Table 5.A1.1, Table 5.A1.2 shows no significant interaction between socio-economic status and linguistic distance. By contrast, the interaction between linguistic distance and gender is statistically significant and strong in Model 4. A 10-point increase on the index of linguistic distance leads to a 1.3 percentage-point reduction in the chance of boys reporting a sense of belonging at school, but only a 0.7 percentage-point reduction in the likelihood of girls reporting so.

Table 5.A1.2 also confirms the critical-period hypothesis, since the main effect of the early-arrival variable is statistically significant in Model 5. However, when also including its interaction with linguistic distance (Model 6), the main effect of the early-arrival variable is not statistically significant, while the main effect of linguistic distance and the interaction of the two variables is significant. For foreign-born students who had arrived at or after the age of 12, a 10-point increase in the linguistic distance between the language they speak at home and the one in which they sat the PISA test reduces their chances of reporting a sense of belonging by 2.4 percentage points. By contrast, among immigrant students who had arrived before the age of 12, a 10-point increase in linguistic distance translates into a 1.7 percentage-point reduction in the likelihood of reporting a sense of belonging at school.

Tables 5.A1.3 and 5.A1.4 present the output from the six models using life satisfaction and anxiety as outcome variables. In Table 5.A1.3, the index of linguistic distance is statistically significant in only three models and at low significance levels. When including GDP and HDI differentials as well as the gender and socio-economic status of students, together with their interaction with linguistic distance in the regression (Models 2b, 3 and 4), a 10-point increase in linguistic distance reduces students' chances of reporting satisfaction with life by 0.4 percentage point. In Table 5.A1.4, linguistic distance is not statistically significant in any of the model specifications except for the fifth, in which it interacts with students' age at arrival and it assumes a positive value. Results show that linguistic distance has an impact on cognitive outcomes and social well-being, but a negligible one on emotional and motivational well-being.

Table 5.A1.5 presents the results from the regressions that were run with motivational resilience as the outcome variable. The coefficient for linguistic distance is not statistically significant, or small and significant at low levels in all models except for the fourth, in which the interaction of the variable with gender is included. In this model, the linguistic distance variable and the interaction with gender are statistically significant. According to Model 4, a 10-point increase in linguistic distance reduces the likelihood of attaining high levels of achievement motivation by 0.3 percentage point among boys, but it increases the chance by 0.02 percentage point among girls. Results indicate that there are significant gender-based differences in the effect of speaking a different language on achievement motivation.

Annex 5.A1

This section explains the models that were used to investigate the effects of linguistic distance on students' academic and well-being outcomes and presents results. For each outcome, a set of eight separate models was developed. First, the outcome of interest was considered to be a function of whether the student has an immigrant background, as well as the linguistic distance between the language spoken at home by the student and the one in which he or she sat the assessment (Model 1). Model 2a is identical to the first but also accounts for potential socio-economic and gender differences. Model 2b is identical to 2a but it controls for differences in GDP and the Human Development Index (HDI) between the host and origin countries of immigrant students in 2010, as do all the following models. (For a detailed explanation of how these measures were obtained, please see Annex 2.) Models 3 and 4 also include interaction terms of linguistic distance with socio-economic background and gender, respectively. Model 5 introduces a binary variable that identifies students who arrived in the host country before the age of 12 and those who arrived later. Model 6 also includes an interaction of that variable with the linguistic-distance indicator. Since the binary variable for early arrival assumes non-missing values only for foreign-born students, the immigrant variable is not included in the last two models. In all tables, the coefficient for linguistic distance reports the effects of a 10-point increase in the linguistic-distance indicator.

Table 5.A1.1 ▪ **Attaining baseline academic proficiency and linguistic distance**

Variable	Model 1	Model 2a	Model 2b	Model 3	Model 4	Model 5	Model 6
Students with an immigrant background	-4.24***	-2.63***	-4.3***	-4.29***	-4.3***		
	-(0.71)	-(0.71)	-(1.18)	-(1.18)	-(1.18)		
Linguistic Distance	-1.62***	-0.74***	-1.08***	-1.15***	-1.28***	-1.23***	-0.74
	(0.20)	(0.19)	(0.20)	(0.21)	(0.22)	(0.47)	(0.70)
Socio-economic background (ESCS)		12.68***	12.85***	12.89***	12.85***	10.07***	10.23***
		(0.20)	(0.19)	(0.20)	(0.19)	(1.38)	(1.39)
Female Students		1.44***	1.64***	1.64***	1.55***	5.15**	5.29**
		(0.34)	(0.38)	(0.38)	(0.38)	(2.18)	(2.20)
ESCS*Linguistic distance				-0.16**			
				(0.08)			
Female students*Linguistic distance					0.4**		
					(0.16)		
Arrivaed before the age of 12						14.31***	14.87***
						(3.52)	(3.81)
Arrived before the age of 12*Linguistic distance							-0.71
							-0.78
Constant	58.5***	63.38***	62.97***	62.99***	63.01***	44.5***	43.91***
	(0.40)	(0.38)	(0.38)	(0.38)	(3.93)	(3.94)	-(3.94)
Observations	445 912	443 802	389 873	389 873	389 873	12 589	12 589
Adjusted R-squared	0.186	0.251	0.264	0.264	0.264	0.294	0.295
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
GDP and HDI differential	No	No	Yes	Yes	Yes	Yes	Yes

Notes: Standard error in parentheses.
* p<0.10, ** p<0.05, *** p<0.01
Source: OECD, PISA 2015 Database and Wichmann, Søren, Eric W. Holman, and Cecil H. Brown (eds.), 2016. The ASJP Database (version 17).
StatLink http://dx.doi.org/10.1787/888933684447

Table 5.A1.2 ▪ **Reporting a sense of belonging at school and linguistic distance**

Variable	Model 1	Model 2a	Model 2b	Model 3	Model 4	Model 5	Model 6
Students with an immigrant background	-2.26***	-1.93***	-3.41***	-3.41***	-3.41***		
	(0.70)	(0.70)	(0.83)	(0.83)	(0.83)		
Linguistic Distance	-0.34**	-0.13	-0.99***	-1.04***	-1.25***	-1.11**	-2.38***
	(0.15)	(0.15)	(0.17)	(0.17)	(0.23)	(0.44)	(0.65)
Socio-economic background (ESCS)		3.03***	3.12***	3.15***	3.13***	2.76**	2.84***
		(0.16)	(0.17)	(0.17)	(0.17)	(1.11)	(1.10)
Female Students		2.77***	3.03***	3.03***	2.92***	3.47	3.56
		(0.34)	(0.31)	(0.31)	(0.31)	(2.32)	-(2.32)
ESCS*Linguistic distance				-0.12			
				(0.08)			
Female students*Linguistic distance					0.52**		
					(0.21)		
Arrivaed before the age of 12						7.34***	5.06
						(2.82)	(3.19)
Arrived before the age of 12*Linguistic distance							1.68**
							-0.71
Constant	66.87***	66.79***	67.3***	67.31***	67.35***	53.28***	55.19***
	(0.26)	(0.34)	(0.32)	(0.32)	(0.32)	(2.84)	(3.15)
Observations	426 121	424 499	373 237	373 237	373 237	12 013	12 013
Adjusted R-squared	0.05	0.055	0.056	0.057	0.057	0.037	0.039
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
GDP and HDI differential	No	No	Yes	Yes	Yes	Yes	Yes

Notes: Standard error in parentheses.
* p<0.10, ** p<0.05, *** p<0.01
Source: OECD, PISA 2015 Database and Wichmann, Søren, Eric W. Holman, and Cecil H. Brown (eds.), 2016. The ASJP Database (version 17).
StatLink http://dx.doi.org/10.1787/888933684466

Table 5.A1.3 ▪ Reporting being satisfied with life and linguistic distance

Variable	Model 1	Model 2a	Model 2b	Model 3	Model 4	Model 5	Model 6
Students with an immigrant background	-4.58*** (0.67)	-4.03*** (0.70)	-5.95*** (1.30)	-5.95*** (1.30)	-5.95*** (1.29)		
Linguistic Distance	-0.13 (0.12)	0.09 (0.12)	-0.4** (0.18)	-0.38** (0.17)	-0.41* (0.23)	-0.8 (0.55)	-0.81 (0.65)
Socio-economic background (ESCS)		3.04*** (0.19)	3*** (0.20)	2.99*** (0.21)	3*** (0.20)	3.7*** (1.07)	3.7*** -(1.07)
Female Student		-7.27*** (0.33)	-6.85*** (0.32)	-6.85*** (0.32)	-6.85*** (0.31)	-5.82*** (2.23)	-5.82*** -(2.23)
ESCS*Linguistic distance				0.06 (0.08)			
Female students*Linguistic distance					0.01 (0.23)		
Arrivaed before the age of 12						1.71 (2.82)	1.69 (3.19)
Arrived before the age of 12*Linguistic distance							0.01 -0.85
Constant	70.85*** (0.23)	75.57*** (0.29)	75.37*** (0.26)	75.36*** (0.26)	75.37*** (0.25)	69.45*** (3.27)	69.47*** -(3.58)
Observations	303 018	301 804	267 673	267 673	267 673	9 827	9 827
Adjusted R-squared	0.037	0.047	0.050	0.050	0.050	0.037	0.037
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
GDP and HDI differential	No	No	Yes	Yes	Yes	Yes	Yes

Notes: Standard error in parentheses.
* p<0.10, ** p<0.05, *** p<0.01
Source: OECD, PISA 2015 Database and Wichmann, Søren, Eric W. Holman, and Cecil H. Brown (eds.), 2016. The ASJP Database (version 17).
StatLink http://dx.doi.org/10.1787/888933684485

Table 5.A1.4 ▪ Reporting low levels of schoolwork-related anxiety and linguistic distance

Variable	Model 1	Model 2a	Model 2b	Model 3	Model 4	Model 5	Model 6
Students with an immigrant background	-2.51*** (0.54)	-2.03*** (0.54)	-2.24*** (0.85)	-2.23*** (0.85)	-2.24*** (0.85)		
Linguistic Distance	-0.11 (0.11)	0.06 (0.11)	0.07 (0.11)	0.05 (0.12)	-0.07 (0.13)	1.01** (0.43)	0.63 (0.65)
Socio-economic background (ESCS)		2.43*** (0.16)	2.31*** (0.19)	2.32*** (0.19)	2.31*** (0.19)	2.85** (1.32)	2.88** (1.32)
Female Student		-13.15*** (0.32)	-13*** (0.35)	-13*** (0.35)	-13.07*** (0.35)	-14.46*** (2.05)	-14.43*** -(2.05)
ESCS*Linguistic distance				-0.04 (0.07)			
Female students*Linguistic distance					0.28 (0.21)		
Arrivaed before the age of 12						5.9** (2.82)	5.18* (3.19)
Arrived before the age of 12*Linguistic distance							0.51 -0.71
Constant	30.08*** (0.21)	37.45*** (0.28)	37.24*** (0.28)	37.24*** (0.28)	37.27*** (0.28)	30.98*** (3.03)	31.59*** (3.16)
Observations	369 321	367 764	320 721	320 721	320 721	12 287	12 287
Adjusted R-squared	0.055	0.079	0.081	0.081	0.081	0.078	0.078
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
GDP and HDI differential	No	No	Yes	Yes	Yes	Yes	Yes

Notes: Standard error in parentheses.
* p<0.10, ** p<0.05, *** p<0.01
Source: OECD, PISA 2015 Database and Wichmann, Søren, Eric W. Holman, and Cecil H. Brown (eds.), 2016. The ASJP Database (version 17).
StatLink http://dx.doi.org/10.1787/888933684504

Table 5.A1.5 ▪ High achievement motivation and linguistic distance

Variable	Model 1	Model 2a	Model 2b	Model 3	Model 4	Model 5	Model 6
Students with an immigrant background	2.01*** (0.45)	2.25*** (0.45)	-0.05 (1.01)	-0.04 (1.01)	-0.05 (1.01)		
Linguistic Distance	-0.12* (0.06)	-0.01 (0.06)	-0.16* (0.09)	-0.2** (0.09)	-0.34*** (0.13)	0.25 (0.28)	0.04 (0.65)
Socio-economic background (ESCS)		1.73*** (0.11)	1.87*** (0.12)	1.9*** (0.12)	1.88*** (0.12)	0.91 (0.66)	0.92 -(0.66)
Female Students		-3.23*** (0.24)	-3.51*** (0.26)	-3.51*** (0.26)	-3.59*** (0.27)	-1.73 (1.65)	-1.71 (1.65)
ESCS*Linguistic distance				-0.1** (0.05)			
Female students*Linguistic distance					0.36** (0.14)		
Arrivaed before the age of 12						3.86 (2.82)	3.47 (3.19)
Arrived before the age of 12*Linguistic distance							0.28 -0.6
Constant	75.94*** (0.18)	78.14*** (0.22)	78*** (0.22)	78.01*** (0.22)	78.04*** (0.22)	76.43*** (2.56)	76.75*** (2.80)
Observations	368 769	367 221	320 206	320 206	320 206	12 285	12 285
Adjusted R-squared	0.187	0.189	0.192	0.192	0.192	0.12	0.12
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
GDP and HDI differential	No	No	Yes	Yes	Yes	Yes	Yes

Notes: Standard error in parentheses.
* p<0.10, ** p<0.05, *** p<0.01
Source: OECD, PISA 2015 Database and Wichmann, Søren, Eric W. Holman, and Cecil H. Brown (eds.), 2016. The ASJP Database (version 17).
StatLink http://dx.doi.org/10.1787/888933684523

Annex 5.A2

This section outlines some of the methodologic issues encountered and the solutions adopted for the final part of the chapter on linguistic distance.

Classification of languages

Although most languages are classified identically within PISA and the Max Planck's ASJP, several do not correspond perfectly. In order to obtain measures of linguistic distance for the greatest possible number of language pairs, certain minor assumptions had to be made, some of which have also been adopted by other researchers using the ASJP and PISA data (see Isphording et al., 2015). Table 5.A2.1 lists the arbitrary conversions that were made.

Table 5.A2.1 ▪ **Conversion of ambiguous cases between PISA and ASJP**

PISA classification	ASJP equivalent adopted
Chinese	Mandarin
Creole (un Spain)	Average of central American creole variants in ASJP
Flemish dialect (BEL)	Belgian
German (LIE)	German
Greenlandic	East Greenlandic
Kurdish	Kurmanji Kurdish
Norwegian	Bokmal
Pushto	Average of ASJP variants
Quechua	Average of variants in ASJP
Raeto'-Romance (in Germany)	Average of Romansh dialects in ASJP
Raeto'-Romance (in Italy)	Friulian
Romani (in Czechoslovakia)	East Slovak Romani
Romani (in Slovenia)	Vend Romani
Romani (in Sweden)	Average of available Romani variants in ASJP
Romani (in the Slovak Republic)	Slovak version of Romani
Serb (Yekavian)	Serbo-croatian
Serbian	Serbo-croatian
Swiss German	German
Swiss Italian	Italian
Yugoslavian – Serbian, Croatian, etc.	Serbo-croatian

Source: OECD, PISA 2015 Database and Wichmann, Søren, Eric W. Holman, and Cecil H. Brown (eds.), 2016. The ASJP Database (version 17).
StatLink http://dx.doi.org/10.1787/888933684542

For several students, the linguistic distance indicator could not be calculated because some languages are classified into broad categories in PISA, such as "another language". Overall, the measure of linguistic distance produced is missing for about 11% of the PISA 2015 sample.

Difference between the GDP and HDI of countries of origin and destination

Data on per capita GDP for PISA-participating countries and on parents' countries of origin were obtained from the World Bank online database. Specifically, data from 2010 at 2011 PPP-adjusted constant prices were used because they covered the largest portion of the sample of PISA destination and origin countries. Data for the Human Development Index were obtained from the United Nations Human Development Programme online database. For the sake of consistency, figures from 2010 were also used for the HDI.

Using HDI and per capita GDP data for PISA countries and parents' countries of origin, differentials between origin and destination countries in HDI and per capita GDP were constructed for each parent in the sample (differentials were zero for native-born parents). Then, each student was assigned the largest differential (either the mother's or the father's differential), such that native students had a value of zero for both the HDI and the GDP differentials.

Notes

1. This classification of a native student applies only to this chapter; in other chapters, native students are also native-born students of native-born parents who do not speak the language of assessment.

2. In some countries and contexts, multiple language communities coexist and students are taught in more than one language. In countries where this is the case, the PISA assessment is available in the relevant multiple languages.

3. Students who reported that they "agree" or "strongly agree" with the statement "I feel like I belong at school" and "disagree" or "strongly disagree" with the statement "I feel like an outsider at school".

4. Students who reported a life satisfaction of 7 or above on a scale from 0 to 10.

5. Students who reported that they "disagree" or "strongly disagree" with the statements "I often worry that it will be difficult for me taking a test" and "Even if I am well prepared for a test, I feel very anxious".

References

Bermudez, A. (1994), *Doing our Homework: How Schools can engage Hispanic Communities*, Appalachian Educational Laboratory, Inc, Charleston, WV.

Bakker, D. et al. (2009), "Adding typology to lexicostatistics: a combined approach to language classification", *Linguistic Typology,* Vol. 12, pp. 167-179.

Bilgili, Ö., T. Huddleston and **A.L. Joki** (2015), *The Dynamics between Integration Policies and Outcomes: a Synthesis of the Literature*, MIPEX Project, Brussels.

Boykin, A.W., K.M. Tyler and **O. Miller** (2005), "In search of cultural themes and their expressions in the dynamics of classroom life", *Urban Education*, Vol. 40/5, pp. 521-549.

Brown, C.H. et al. (2008), "Automated classification of the world's languages: A description of the method and preliminary results", *STUF-Language Typology and Universals Sprachtypologie und Universalienforschung*, Vol. 61/4, pp. 285-308.

Capps, R. (2005), *The Health and Well-Being of Young Children of Immigrants*, Urban Institute, Washington, D.C.

Chiswick, B.R. and **P.W. Miller** (1996), "Ethnic networks and language proficiency among immigrants", *Journal of Population Economics*, Vol. 9/1, pp. 19-35.

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

Clarke, A. and **I. Isphording** (2017), "Language barriers and immigrants health", *Health Economics*, Vol. 26/6, pp.765-778.

Coll, C.G. and **K. Magnuson** (1997), "The psychological experience of immigration: A developmental perspective", in Booth, A., A.C. Crouter and N.S. Landale (eds.), *Immigration and the Family: Research and Policy on U.S. immigrants*, Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 91-131.

Council of the European Union (2004), *Immigrant Integration Policy in the European Union,* Brussels, 19 November, 14615/04 (Presse 321).

Dawson, B.A. and **S.A. Williams** (2008), "The impact of language status as an acculturative stressor on internalizing and externalizing behaviors among Latino/a children: A longitudinal analysis from school entry through third grade", *Journal of Youth and Adolescence*, Vol. 37/4, pp. 399-411.

Dustmann, C. and **A. van Soest** (2001), "Language fluency and earnings: Estimation with misclassified language indicators", *The Review of Economics and Statistics,* Vol. 83/1, pp. 663-674.

Dustmann, C. and **F. Fabbri** (2003), "Language proficiency and labour market performance of immigrants in the UK", *The Economic Journal,* Vol.113/489, pp. 695-717.

Dustmann, C., S. Machin and **U. Schönberg** (2010), "Ethnicity and educational achievement in compulsory schooling", *The Economic Journal*, Vol. 120/546, pp. F272-F297.

Espenshade, T.J. and **C.A. Calhoun** (1993), "An analysis of public opinion toward undocumented immigration", *Population Research and Policy Review*, Vol. 12/3, pp. 189-224.

Frijters, P., Shields, M.A. and **S.W. Price** (2005), "Job search methods and their success: A comparison of immigrants and natives in the UK", *The Economic Journal*, Vol. 115/507.

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

Gil, A.G., W.A. Vega and **J.M. Dimas** (1994), "Acculturative stress and personal adjustment among Hispanic adolescent boys", *Journal of Community Psychology*, Vol. 22/1, pp. 43-54.

Gordon, M.M. (1964), *Assimilation in American Life: The Role of Race, Religion, and National Origins,* Oxford University Press on Demand, Oxford.

Hayfron, J.E. (2001), "Language training, language proficiency and earnings of immigrants in Norway." *Applied Economics,* Vol. 33/15, pp. 1971-1979.

Isphording, I.E. (2015), "What drives the language proficiency of immigrants?", IZA World of Labor 2015: 177.

Isphording, I.E. and **S. Otten** (2013), "The costs of Babylon – linguistic distance in applied economics", *Review of International Economics*, Vol. 21/2, pp. 354-369.

Isphording, I.E. and **S. Otten** (2014), "Linguistic barriers in the destination language acquisition of immigrants", *Journal of Economic Behavior and Organization*, Vol. 105, pp. 30-50.

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

Joppke, C. (2007), "Beyond national models: Civic integration policies for immigrants in Western Europe", *West European Politics*, Vol. 30/1, pp. 1-22.

Leslie, D. and **J. Lindley** (2001), "The impact of language ability on employment and earnings of Britain's ethnic communities", *Economica*, Vol. 68/272, pp. 587-606.

Levenshtein, V.I. (1966), "Binary codes capable of correcting deletions, insertions, and reversals", *Soviet Physics Doklady,* Vol. 10/8, pp. 707-710.

Moles, O. (1993), "Collaboration between schools and disadvantaged parents: Obstacles and openings", in Chavkin, N.F. (ed.), *Families and Schools in a Pluralistic Society*, SUNY Press, Albany, NY.Newport, E.L. (2002), "Critical periods in language development", in Nadel, L. (ed.), *Encyclopedia of Cognitive Science,* Macmillan Publishers Ltd./Nature Publishing Group, London, pp. 737-740.

OECD (2015a), *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence, PISA*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264229945-en.

OECD (2015b), *Immigrant Students at School: Easing the Journey towards Integration*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264249509-en.

OECD (2010), *Closing the Gap for Immigrant Students: Policies, Practice and Performance, OECD Reviews of Migrant Education*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264075788-en.

OECD (2006), *Where Immigrant Students Succeed: A comparative Review of Performance and Engagement in PISA 2003,* OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264249509-en.

Padilla, A.M. and **W. Perez** (2003), "Acculturation, social identity, and social cognition: A new perspective", *Hispanic Journal of Behavioral Sciences*, Vol. 25/1, pp. 35-55.

Romero, A.J. and **R.E. Roberts** (2003), "Stress within a bicultural context for adolescents of Mexican descent", *Cultural Diversity and Ethnic Minority Psychology*, Vol. 9/2, p. 171.

Smart, J. F. and **D.W. Smart** (1995), "Acculturative stress of Hispanics: Loss and challenge", *Journal of Counseling and Development, JCD*, Vol. 73/4, p. 390.

Van Tubergen, F. and **M. Kalmijn** (2005), "Destination-Language Proficiency in Cross-National Perspective: A Study of Immigrant Groups in Nine Western Countries", *American Journal of Sociology*, Vol. 110/5, pp. 1412-1457.

Van Tubergen, F., I. Maas and **H. Flap** (2004), "The economic incorporation of immigrants in 18 Western societies: Origin, destination, and community effects", *American Sociological Review*, Vol. 69/5, pp. 704-727.

Wichmann, S., Er.W. Holman and **C.H. Brown** (eds.) (2016), The ASJP Database (version 17) available on line, http://asjp.clld.org/ (accessed on 18 February 2018).

Zhou, M. and **Y.S. Xiong** (2005), "The multifaceted American experiences of the children of Asian immigrants: Lessons for segmented assimilation", *Ethnic and Racial Studies*, Vol. 28/6, pp. 1119-1152.