IFAU
Institute for Evaluation of Labour Market and Education Policy

Relations between immigration and adult skills:
Findings based on PIAAC

Patrik Lind
Erik Mellander

WORKING PAPER 2016:21

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala. IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

More information about IFAU and the institute's publications can be found on the website www.ifau.se

ISSN 1651-1166

Relations between immigration and adult skills: Findings based on PIAAC[a]

by

Patrik Lind[b] and Erik Mellander[c]

November 23, 2016

Abstract

The international survey of adult skills, PIAAC, records large differences in numeracy and literacy skills between immigrants and non-immigrants. We examine how these differences relate to the countries' average skills and skill rankings. Immigrants are defined by country of birth or in terms of languages spoken. For almost all countries, the differences in average skills between non-immigrants and the country's entire population are significant but small. Regarding skill rankings significant differences are found only for Sweden and these are found to be sensitive to the treatment of individuals that could not conduct the skill tests due to language difficulties.

Keywords: PIAAC, migration, language skills, average scores, rank uncertainty
JEL-codes: I24, J15, O15, C15

[a] Many thanks for helpful comments from Anders Stenberg, Adrian Adermon, George Leckie, Martin Lundin and seminar participants at IFAU. Funding from NordForsk (Grant No. 54681) is gratefully acknowledged

[b] Institute for Evaluation of Labour Market and Education Policy, patrik.lind@ifau.uu.se
[c] Institute for Evaluation of Labour Market and Education Policy, erik.mellander@ifau.uu.se

Table of contents

1 Introduction

In the OECD survey of adult skills, PIAAC[1], it is noted that in several countries the differences in skills between natives and immigrants are quite large to the disadvantage of the latter, cf. OECD (2013a, pp. 125-132). Furthermore, the population shares of immigrants in some of these countries are substantial, too; see, e.g., OECD (op. cit., Table B3.10). This raises a question concerning the interpretation of the reported mean country skill scores.

It is already well known that cross-country differences with respect to culture, language, and institutions may invalidate the interpretation of a country's mean score in an international skill survey as an indicator of the quality of the country's educational and training systems; cf., e.g., Blum et al. (2001) and Bonnet (2002). The issue of immigration adds another dimension to this issue: if a large part of a country's population is made up of low-skilled immigrants, many of whom have obtained their schooling abroad, this may introduce a downward bias in the country's mean score as a measure of how successful the country has been in educating and training its citizens. In this paper we, therefore, consider whether immigrants mattered for the mean scores of the countries that participated in PIAAC, in an absolute and/or relative sense.

With an absolute influence we mean that the country's mean score with respect to its (entire) population differs from its mean score for non-immigrants only. If so, this tells us that the country's skill difference between immigrants and natives is substantial *and* that the immigrants constitute a non-negligible share of the country's population. We test whether the difference between the mean scores of the country's non-immigrants and its entire population is statistically significant. Moreover, we assess its magnitude by relating it to the mean score of the country's entire population.

However, many would probably be more interested knowing whether the immigrants matter with respect to the country's position in the international skill distribution, i.e. in a relative sense. This might be the case even if the country's immigrants do not matter in an absolute sense. And, conversely, a significant absolute difference does not imply a significant relative difference. In general, the latter will be determined by the conditions in other countries to a larger extent than by domestic circumstances.

[1] Programme for International Assessment of Adult Competencies.

To measure if immigrants matter for countries' skill scores in a relative sense, we construct non-immigrant country rankings, i.e. rankings of average skills when immigrants are excluded. We then compare these rankings to the rankings based on the countries' overall averages, i.e. the rankings provided in OECD (2013a). In so doing, we account for the statistical uncertainty associated with country rankings.

Hitherto we have argued as if immigrants constitute a homogeneous group. Of course, this is not the case. Using PISA data for Denmark, matched with register data, Rangvid (2010) shows that the measured negative effects on the host country's skills may vary with the immigrants' country of origin, not only upon entry but also regarding development over time. Several papers have also addressed the importance of the immigration policy pursued by the recipient country. Also using PISA 2000, Entorf and Minou (2005), show that while immigrant skills are below the skills of natives (except in Canada), immigrant skills are higher in what they call 'traditional immigration countries', like Australia, Canada and New Zealand (but not in the USA), than in, e.g., Germany and the Scandinavian countries. Kahn (2004) and Korpi (2012) corroborate these findings with respect to adult skills, as measured in IALS[2]. Kahn (op.cit.) further makes the interesting observation that whereas the skill distributions of natives are single-peaked, the skill distributions of immigrants are bi-modal. Thus, for some countries immigration may actually contribute positively to the skills of its population and how immigrants influence a country's skill will depend on for how long they have lived in the country. Accordingly, we consider all of the countries that participated in PIAAC.

Further, it is not obvious that country of birth is the most appropriate dimension along which to compare the skills of immigrants and natives. Consider, e.g., early-age immigrants. Many of them are likely to have attended the same school system as the natives. To account for this aspect, immigration status may be based on the first or second language spoken by the participants. This alternative is especially relevant in view of the fact that the PIAAC survey was only conducted in the participating countries' official language(s). Thus, we employ definitions of immigrants based on languages as well as country of birth.

[2] International Adult Literacy Survey.

Our paper contains four contributions. The first contribution is that we consider how immigration relates to two of the three different skills measured in PIAAC, namely literacy skills and numeracy skills.[3] The second contribution is that we investigate how robust our results are to alternative definitions of immigrants (and non-immigrants). Specifically, we compare definitions based on country of birth and spoken languages, respectively, and check for common factors that might be decisive for both of the definitions. The third contribution is that we show, by reference to earlier literature, that the conventional confidence intervals used in OECD (2013a) are inappropriate measures of rank uncertainty and, moreover, under-estimate the true uncertainties, compared to an appropriate measure suggested by Leckie and Goldstein (2011). Partly, this result derives from the well-known fact that if a given data set is subjected to a large number of tests the likelihood of rejecting at least one of the (many) null hypotheses is larger than the likelihood of rejection in the context of a single test.[4] The fourth contribution is that we apply the Leckie and Goldstein (op.cit.) measure of rank uncertainty to examine to what extent our estimates of immigration-related changes in the countries' average skills matter for country skill rankings. To this end, we compare country skill rankings corresponding to non-immigrants (only) to the rankings based on the countries' entire populations.

For brevity, and to facilitate comparisons with earlier studies, we focus on numeracy skills in the main text, while the findings with respect to literacy are only commented

[3] In principle, it would be possible to conduct a corresponding analysis also with respect to third skill measured in PIAAC, problem solving in technology-rich environments (PS-TRE). However, for three reasons, we have decided not to do so. The first reason is that the assessment of PS-TRE skills was optional and four countries abstained from it. Secondly, in the countries that did test PS-TRE skills many of the respondents were unable to complete the test because they could not, or did not want to, use a computer. And the corresponding non-response rates varied widely across countries – from 12.1 percent in Sweden to 49.8 percent in Poland – making comparisons across countries difficult to interpret. Due to this non-response problem OECD recommends that a country's PS-TRE skills should not be expressed in terms of test score points, like the literacy and numeracy skills, but in terms of the share of its population that can handle PS-TRE tasks that "… typically require the use of both generic and more specific technology applications." (OECD, 2013a, p. 88). The resulting difference in test outcome measures between literacy and numeracy skills, on the one hand, and PS-TRE skills, on the other hand, provides us with the third reason to limit our attention in this paper to literacy and numeracy skills.

[4] One way to account for multiple tests is to extend the conventional confidence intervals by means of the Bonferroni correction, cf. Dunn (1961). However, that procedure instead entails the risk of *over-estimating* the rank uncertainty because it presupposes that the conventional confidence intervals are correct when (only) two countries are compared. Healy and Goldstein (1995) show that this need not be the case. In particular, they demonstrate that when the standard deviations of two estimated averages are equal the corresponding conventional confidence intervals are much wider than the true intervals; the difference decreases, the more unequal the standard deviations are. In our context the standard deviations of the countries' average skill scores are quite different, implying that the potential over-estimation of rank uncertainty is dominated by under-estimation resulting from negligence of multiple testing (i.e. multiple country comparisons). Like the Bonferroni correction, the Leckie and Goldstein (2011) method accounts for multiple testing, but avoids the over-estimation problem associated with the Bonferroni correction.

upon very briefly. However, tables and figures providing details about the results for literacy skills can be found in Appendix 1.

In summary, our results are the following. For almost all countries participating in PIAAC, average skills are statistically significantly higher when immigrants are excluded compared to when the country's entire population is considered. However, in relative terms the differences are not very substantial; the largest ones amount to 3.5 percent. This holds true irrespective of whether immigrant status is defined in terms of country of birth or in terms of languages spoken.

Country rankings based on non-immigrants only are found to be very similar to the country rankings based on the countries' entire populations. Of the 21 countries considered only two show statistically significant changes in rankings. And of these, only one country – Sweden – changes its rank substantially when non-immigrants only are considered; for numeracy skills Sweden's rank is improved by several positions, under both of the two immigrant definitions employed.[5] However, these results are found to be sensitive with respect to the treatment of individuals that could not take the skill assessment tests due to language difficulties.

The paper unfolds as follows. The next section contains a description of the PIAAC survey which focuses on how the survey was conducted, its informational content, sample size considerations and the treatment of language-related non-response. Section 3 describes the definitions of immigrants based on country of birth and spoken languages that we consider. Changes in average skills associated with immigration are considered in Section 4 while Section 5 deals with the implied influences on country skill rankings. Section 6 reports on the sensitivity analysis of the Swedish rank results mentioned in the previous paragraph. Section 7 concludes.

2 The PIAAC survey[6]

The PIAAC survey was initiated by the OECD and conducted 2011-2012.[7] Altogether, 166 000 individuals in 24 different countries participated in the survey. The target population was the non-institutionalized population aged 16-65, residing in the country

---

[5] The other country is France, which improves its rank in literacy by one position, under of the immigrant definitions.

[6] This section is based on three sources: OECD (2013a, the Overview chapter), OECD (2013b, Chapter 3), and OECD (2013c).

[7] During 2014-2015 a second wave of PIAAC was conducted, involving nine additional countries, cf. OECD (2016). However, this paper is based only on data from the first, 2011-2012, wave of PIAAC.

at the time of data collection, irrespective of nationality, citizenship, or language status. The 24 participating countries/parts of countries were: Austria, Australia, Belgium (Flanders), Canada, Cyprus, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Ireland, Italy, Japan, Korea, the Netherlands, Norway, Poland, the Russian Federation, the Slovak Republic, Spain, Sweden, UK (England and Northern Ireland), and the USA.[8]

2.1 Implementation, informational content, and sample sizes

The PIAAC survey was conducted in the respondent's home or in another location agreed upon by the respondent and the interviewer. It was administered in two stages. The first stage was an extensive Background Questionnaire (BQ), resulting in 258 variables measuring demographic characteristics, educational and labor market experiences, and activities related to the skills assessed. In the second stage cognitive skills were assessed in three domains: literacy, numeracy, and problem solving in technology-rich environments (PS-TRE). These are defined as follows (OECD, 2013a):

- **Literacy**: The ability to understand, evaluate, use and engage with *written* texts to participate in society, to achieve one's goals and to develop one's potential.
- **Numeracy**: The ability to access, use, interpret and communicate mathematical information and ideas in order to engage in and manage the mathematical demands of a range of situations in adult life.
- **Problem solving in technology-rich environments (PS-TRE)**: The ability to use digital technology, communication tools and networks to acquire and evaluate information, communicate with others and perform practical tasks.

Assessment of PS-TRE was optional and four countries – Cyprus, France, Italy, and Spain – opted out.

The BQ was administered on the interviewer's computer. Respondents were able to seek assistance from others when having difficulties to answer, due to, e.g., language problems. The cognitive assessments were primarily conducted on the interviewer's computer and to a lesser extent (about 25 percent) by completing printed test booklets, for respondents unable or unwilling to use a computer. When taking the cognitive assessment tests, respondents were not permitted to seek assistance from others. The

[8] The Russian Federation is not included in the following, as the data from the Russian Federation are preliminary and do not include the population of the Moscow municipal area.

background questionnaire took between 30 and 45 minutes to complete, while the time used for the cognitive assessment varied between 40 and 50 minutes.

The sample size requirements determined by the OECD depended on the number of cognitive domains assessed and the number of languages in which the assessment was administered. For countries administering PIAAC in one language and assessing only literacy and numeracy the minimum sample size was 4 500. If, in addition, PS-TRE skill were assessed the minimum increased to 5 000. Minimum sample sizes were further increased for countries administering PIAAC in multiple languages, the increases depending on the shares of the population speaking the different languages and if the country's survey results were to be published in more than one language.

PIAAC was designed to provide accurate estimates of proficiency across the adult population and major subgroups of it, rather than for individuals. This made it possible to reduce the time required for the cognitive assessment by implementing the survey such that each respondent was only given a subset of the test items; no individual took tests in all of the three skill domains assessed, neither was (s)he confronted with all of the test items within any of the skill domains (76 items in literacy, 76 in numeracy and 14 in PS-TRE). The resulting uncertainty with respect to the individual's skills was handled by representing her/his proficiency by means of a set of 10 'plausible values' drawn from a response distribution estimated by means of test item responses and BQ variables, for groups of similar individuals. Separate individual response distributions are estimated for each of the three skill domains.

Thus, in addition to the sampling error that is intrinsic to all survey data, the PIAAC results also are characterized by uncertainty in the individual skill estimates. Both of these stochastic elements are accounted for in our empirical analyses, by means of re-sampling procedures.[9]

Table 1 provides information about skills assessed, assessment language(s), minimum sample requirements, and achieved samples, together with the corresponding response rates, for each of the participating countries. It can be seen in the table that five countries administered PIAAC in multiple languages: Canada, Estonia, Finland, the Slovak Republic, and Spain. It is noteworthy that in consequence of the fact that the minimum sample sizes were independent of the countries' population sizes, the USA,

[9] In accordance with the procedures specified and adopted by the OECD. For details, see Appendix 2: Computation of variances allowing for both sampling error and skill measure uncertainty

with 322 million inhabitants, was subject to the same minimum sample size as, e.g., Norway, with a population of 5 million. This means that in terms of sample sizes alone, potential problems with respect to representativity and validity are likely to be more severe in countries with large populations.

According to Table 1, Sweden was the only country (beside the Russian Federation) for which the size of achieved sample was below the minimum required sample size. Moreover, Sweden had the lowest response rate of all the participating countries. Nevertheless, the Swedish results were included in OECD (2013a), because the Swedish sample was judged to be representative of its population and because Sweden made large efforts to reduce the non-response bias in its results.[10]

[10] Sample representativity is judged by, i.a., the percentage of the target population not included in the sample; cf. OECD (2013c, Table 16-2). Concerning efforts to reduce non-response bias, see OECD (2013c, Chapter 16.3).

Table 1: Skills assessed, assessment language(s), and sample sizes, and response rates[1]

Country	Cognitive skills assessed [2]	Assessment language(s)	Minimum sample requirements [3]	Achieved sample	Response rate, % [4]
Australia	L, N, PS-TRE	English	5 000	8 600	71
Austria	L, N, PS-TRE	German	5 000	5 130	53
Belgium (Flanders)	L, N, PS-TRE	Dutch	5 000	5 463	62
Canada	L, N, PS-TRE	English, French	10 000	27 285	59
Cyprus	L, N	Greek	4 500	5 053	73
Czech Republic	L, N, PS-TRE	Czech	5 000	6 102	66
Denmark	L, N, PS-TRE	Danish	5 000	7 328	50
Estonia	L, N, PS-TRE	Estonian, Russian	7 500	7 632	63
Finland	L, N, PS-TRE	Finnish, Swedish	5 276	5 464	66
France	L, N	French	4 500	6 993	67
Germany	L, N, PS-TRE	German	5 000	5 465	55
Ireland	L, N, PS-TRE	English	5 000	5 983	72
Italy	L, N	Italian	4 500	4 621	56
Japan	L, N, PS-TRE	Japanese	5 000	5 278	50
Korea	L, N, PS-TRE	Korean	5 000	6 667	75
Netherlands	L, N, PS-TRE	Dutch	5 000	5 170	51
Norway	L, N, PS-TRE	Norwegian	5 000	5 128	62
Poland	L, N, PS-TRE	Polish	5 000	9 366	56
Slovak Republic	L, N, PS-TRE	Slovak, Hungarian	5 568	5 723	66
Spain	L, N	Castilian, Basque, Catalan, Galician, Valencian	6 000	6 055	48
Sweden	L, N, PS-TRE	Swedish	5 000	4 469	45
UK (England & Northern Ireland)	L, N, PS-TRE	English	5 000	8 892	60
USA	L, N, PS-TRE	English	5 000	5 010	70

[1] The Russian Federation is not included in table as the data from the Russian Federation are preliminary and do not include the population in the Moscow municipal area.
[2] L = Literacy, N = Numeracy, PS-TRE = Problem Solving in Technology-Rich Environments.
[3] The minimum sample requirements accounted for number of skills assessed and for multiple languages. There were no sample size requirements with respect to oversampling, however. Five countries oversampled: Australia – certain states and territories; Canada – 16-25 year olds, provinces/territories, linguistic minorities, aboriginals, and recent immigrants; Czech Republic – 16-29 year olds; Denmark – 55-65 year olds and recent immigrants; Poland – 19-26 year olds.
[4] Computed as the ratio (× 100) between the number of respondents and the number of individuals in the sampling frame. Targeted to be at least 70 %. OECD (2013c, Table 16-4).

Sources: OECD (2013b, Table 3.7), OECD (2013c, Table 14-19, Table 16-7)

2.2 The treatment of language-related non-response on the skill tests

For the analyses in the following sections, it is important to note that the response rates in PIAAC do *not* correspond to persons that have completed the BQ and taken skill assessment tests. With respect to the skill assessment tests, three categories of individuals can be distinguished. The first consists of persons that have taken the tests. Common to the second and third categories is that they are made up of persons that have not taken the skill tests due to language difficulties. However, the persons in the second category have been assigned imputed test scores while the individuals in the third category have not. This difference stems from the extent to which the individuals have completed the BQ.

The persons in the second category have, at least, provided responses to key background questions: age, gender, highest level of schooling, and employment status. Based on their reported characteristics, they have received imputed skill scores for literacy and numeracy, (but not for PS-TRE). As a rule, these imputed scores are very low; cf. (OECD 2013c, Ch. 17).

For persons in the third category there is not information about all key background characteristics. Lacking these data, no attempt has been made to impute literacy and numeracy scores.

Some countries made extensive use of the possibility to seek assistance from others when completing the BQ (cf. above). For example, Sweden hired interpreters that accompanied the interviewers, thus enabling imputation of skill scores also for individuals with low proficiency in the country's language(s) of assessment.

Immigrants are over-represented among the respondents not taking the skill assessment tests. This means that differences across countries with respect to whether or not they have imputed skill scores to individuals not taking the tests can matter for how the results of individual countries relate to immigration. That there are cross-country differences along this dimension is clear from Table 2, below. The first column of Table 2 shows the country shares of respondents in the second category discussed above while the second column shows the shares in the third category.

Table 2: Proportions of respondents not taking skill assessment tests, due to language difficulties; partitioned according to whether or not they have been assigned imputed scores [1,2]

Country	Respondents with imputed scores, % [3]	Respondents without imputed scores, %
Australia	4.9	1.9
Austria	1.5	1.8
Belgium (Flanders)	0.6	5.2
Canada	4.7	0.9
Cyprus	0.2	17.7
Czech Republic	0.3	0.6
Denmark	5.0	0.4
Estonia	1.7	0.4
Finland	6.1	0.0
France	6.5	0.8
Germany	1.7	1.5
Ireland	3.3	0.5
Italy	3.9	0.7
Japan	0.1	1.2
Korea	2.2	0.3
Netherlands	1.7	2.3
Norway	4.6	2.2
Poland	1.1	0.0
Slovak Republic	1.6	0.3
Spain	2.0	0.8
Sweden	5.9	0.0
UK (England & Northern Ireland)	2.5	1.4
USA	2.3	4.2

[1] The Russian Federation is not included in table as the data from the Russian Federation are preliminary and do not include the population in the Moscow municipal area.
[2] The proportions in the table are weighted so as to reflect population shares.
[3] The imputed scores concern literacy and numeracy; no scores were imputed for PS-TRE skills.

Source: OECD (2013b, Table 3.10).

To facilitate the interpretation of Table 2, it is reasonable to assume that, on average, respondents who did not take skill tests have lower skills than respondents who took skill tests.[11] This means that respondents for which skill scores are imputed contribute negatively to their country's average skills while, on average, this is not true for individuals who are left out; these individuals are categorized as literacy-related non-respondents.[12]

In three countries, the treatment of language-related non-response unambiguously induces a decrease in the estimated average skills. These countries are Finland, Poland, and Sweden, in which all the language-related non-respondents have been assigned imputed scores, as indicated by shares equal to zero in the second column of Table 2. At the other extreme, Belgium (Flanders), the USA, and, in particular, Cyprus exhibit large shares of language-related non-respondents for which skill scores have not been imputed; cf., again, the second column of Table 2.[13]

3 Alternative definitions of immigrants

In official statistics, immigrants are defined in terms of country of origin. This is also the definition predominantly used in OECD (2013a) where, however, language-based definitions are employed, too. We consider how defining immigrants as 'foreign-born' or 'non-native speakers' affects the country shares of immigration in the total population, and the average skills and average years of schooling among the immigrants.

In this and the following section not only the Russian Federation will be left out of the analysis but also Australia and Cyprus. The Cypriot data have been withdrawn by the OECD and the use of Australian data requires an approval from the Australian Bureau of Statistics, which we currently do not have.

3.1 Definitions based on country of birth

In PIAAC, the definition of immigrants based on country of birth is the following:

[11] This assumption is in accordance with the fact that, as mentioned above, individuals for which skills scores are imputed receive very low scores.
[12] That individuals for whom skills are not imputed are left out in the computation of the country's average skills effectively means that these individuals are treated as average performers in their respective countries, which is contrary to the assumption that respondents that did not take skill tests have lower skills than respondents who did.
[13] The share for Cyprus, 17.7 percent, is unreasonably high. Presumably, it constitutes (one of) the reason(s) why the OECD has withdrawn the Cypriot data; see, e.g., note 2 to Table 3.

D1: The respondent is classified as immigrant if her/his country of birth is not equal to the country of assessment.

We will, however, use the following, slightly different, definition:

D2: The respondent is classified as immigrant if *D1* applies **and** there is non-missing information for her/him about the first (two) language(s) that (s)he first learned at home during childhood and still understands and the language that (s)he mostly speaks at home.[14]

The reason why we prefer *D2* to *D1* is that the immigrants defined by *D2* constitute a subset of the same set of respondents relative to which we define immigrants in terms of language(s) spoken, namely *D3* and *D4*; cf. Section 3.2. This means that the comparison of immigrants defined by country of birth and spoken language(s), respectively, is not confounded by differences in the underlying groups of respondents.

Neither of the definitions *D1* and *D2* preclude that a respondent classified as immigrant may master (one or several of) the language(s) in which the proficiency tests are conducted. This may be due to, e.g., immigration at an early age, the language of assessment being a language with large international coverage, like English or Spanish, or, in some countries, the use of multiple test languages; cf. Table 1 above.

As one would expect, the immigrant groups defined by *D1* and *D2* are very similar. Essentially, there is only one country for which the distinction between *D1* and *D2* matters, namely Finland. Under *D1*, the share of immigrants in the Finnish population is estimated to be 5.7 percent, while under *D2* the estimated share is 3.4 percent.[15] For the

[14] *Missing* information is recorded for respondents answering that they don't know or refuse to answer altogether. These respondents are excluded from the analysis below.

[15] This difference is due to confidentiality reasons. Part of the information collected in Finland about the first language(s) the respondents learned at home during childhood was aggregated before it was made available in Finland's public use file. The aggregate, comprising several very small groups of individuals which together represent 2.7 percent of the Finnish population, was coded as missing. Our Finnish colleagues have informed us that, from *national* data files, it can be inferred that all of the individuals in the aggregate should be defined as immigrants according to definition *D2*. However, the national data files also tell that some 20 percent of the aggregate's individuals mostly speak Finnish or Swedish at home. The corresponding individuals should *not* be classified as immigrants according to the definition *D4* considered below. The problem is that these individuals cannot be identified in the public use file. Accordingly, the condition requiring non-missing information about the language mostly spoken at home is not satisfied for the aggregate coded as missing and so it has to be left out of the analysis. This substantially reduces the share of immigrants under the definition *D2* as all individuals in the left out aggregate are immigrants according to *D2*.

other countries, the largest of the corresponding differences are around half of a percentage point. [16]

For definition *D2*, Table 3 shows the countries' immigrant shares in the population, the mean numeracy proficiency skills among the immigrants, and their average years of schooling, together with skills and years of schooling for the non-immigrants, i.e. the native-born.

Table 3 shows that the immigrant population share varies markedly across countries: from one fourth in Canada and one fifth in Ireland to very close to zero in Poland and Japan. In the table, the countries have been ordered according to the size of their immigrant population share, since a 'large' immigrant population share is a necessary (but not sufficient) condition for the immigrants to matter for the countries' average skill scores.

With respect to mean numeracy scores it can be noted that for all of the countries in Table 3 the mean immigrant scores are below the OECD average of 269. In addition to (average) scores, the outcomes of the skill assessments are also reported in OECD (2013a) in terms of six skill levels: 'Below level 1', Level 1, Level 2, … , Level 5, where 'Below level 1' corresponds to at most 176 score points, Levels 2-4 are defined by successive addition of score point intervals comprising 50 points, and Level 5 is defined by at least 376 score points, cf. note 3 to Table 3. For two of the countries, Poland and France, the mean immigrant scores correspond to Level 1. For all other countries the mean immigrant scores correspond to proficiency Level 2. At Level 2 of the numeracy scale, respondents are able to identify and act on mathematical information in common contexts that are fairly explicit. The tasks usually requires the application of two or more steps of calculation with whole numbers and with decimals, percentages and fractions, simple measurement and spatial representation, and interpretation of relatively simple information in tables and graphs (OECD, 2013b).

Regarding the differences between the mean scores of immigrants and non-immigrants, it can be deduced from Table 3 that these are all negative, with one exception: in Ireland the mean immigrant score is 0.9 points higher than the mean score

[16] Qualitatively, the results generated under the definition *D2*, which are reported below, are the same as the results generated under definition *D1*, but the point estimates differ for Finland. In particular the average score among immigrants is 26 score points lower. As the mean score among natives does not differ, the native-foreign difference under *D1* is twice the corresponding difference under *D2* (52 compared to 26). Average years of schooling drops from 12.4, under *D2*, to 12.2, under *D1*. The full set of results generated under *D1* can be obtained from the authors, on request.

of native-born.[17] The largest negative differences are found in France (-70.8) and Sweden (-56.2), both of which exceed one skill level (50 score points).

Table 3 also provides information about whether the skill differences to the immigrants' disadvantage are matched by lower (numbers of years of) education. For France, this certainly is the case. In addition to having the lowest average score among immigrants, France also exhibits the largest (negative) difference in years of schooling between immigrants and natives. In general, the relationship between measured skills and schooling does not appear unambiguous, however. For instance, Norway which has the third largest negative difference in mean scores between immigrants and natives, displays a slightly *positive* immigrant-native difference with respect to years of schooling. Computing the correlation between immigrant-native differences in average scores and average years of schooling for all the countries we obtain a coefficient of correlation equal to 0.67.

[17] These differences can be computed from Table 3 and they can also be obtained directly from Table 5.

Table 3: Immigrant and non-immigrant characteristics, and mean numeracy scores when immigrants are defined according to definition D2.[1] Countries ordered in descending order by immigrant population shares

Country [2]	Immigrant population share, %	Immigrant mean numeracy score [3]	Non-immigrant mean numeracy score [3]	Immigrant mean years of schooling	Non-immigrant mean years of schooling
Canada	25.7	249.8	270.6	13.8	12.9
Ireland	21.0	256.3	255.4	15.3	14.3
Sweden	17.5	232.7	288.9	11.7	12.1
Austria	16.3	248.4	280.2	11.7	11.8
UK (England & Northern Ireland)	14.9	238.5	265.9	13.0	12.3
USA	14.7	226.1	257.5	12.6	13.2
Germany	13.8	239.6	276.9	11.9	12.6
Norway	13.4	238.3	284.6	14.1	13.9
Spain	13.2	227.0	248.7	11.2	11.3
Estonia	13.0	259.9	275.3	12.0	12.1
France	12.8	215.6	286.4	10.1	13.2
Netherlands	12.8	239.5	259.9	12.6	11.4
Denmark	11.8	245.6	282.7	12.5	12.5
Italy	9.3	231.6	248.7	10.3	10.5
Belgium (Flanders)	6.9	252.0	283.0	12.0	12.4
Czech Republic	4.0	263.0	276.2	13.1	12.9
Finland	3.4	259.6	285.3	12.4	12.2
Slovak Republic	2.3	267.7	276.0	12.6	12.9
Korea	1.6	230.9	263.9	11.9	12.6
Japan	0.4	263.3	288.3	12.6	12.9
Poland	0.2	222.9	259.9	12.8	12.5

[1] *D2*: The respondent's country of birth is not equal to the country of assessment and there is non-missing information about the first (two) language(s) that (s)he learned at home during childhood and still understands, and the language that (s)he mostly speaks at home.

[2] The countries Australia, Cyprus and the Russian Federation are not included in table. Access to the Australian data requires an approval from the Australian Bureau of Statistics, which we, currently, do not have. The Cypriot data have been withdrawn by the OECD. The data from the Russian Federation are preliminary and do not include the population in the Moscow area.

[3] For reference, the mean numeracy score in PIAAC equals 269 and the proficiency levels for numeracy are defined according to: Below Level 1 < 176, 176 £ Level 1 < 226, 226 £ Level 2 < 276, 276 £ Level 3 < 326, 326 £ Level 4 < 376, and 376 £ Level 5.

Source: Own computations.

Table A1 in the Appendix complements Table 3 by providing the corresponding table for literacy skills. The findings in Table A1 are similar to those in Table 3 but, on average, the differences between immigrants and non-immigrants are somewhat smaller in Table A1. The largest difference is found for Sweden (-53.7 score points), followed by the Netherlands (-42.6 score points).

3.2 Definitions based on primary language(s)

The PIAAC Background Questionnaire contains a question about the first language that the respondent learned at home during childhood and still understands. For respondents that spontaneously mention two languages, instead of one, both languages are recorded. For respondents with non-missing answers on this question the following language-based definition of immigrants is used in PIAAC:

D3: The respondent is classified as immigrant if the first (or second) language that (s)he learned at home during childhood, and still understands, is not equal to (one of) the language(s) of assessment.

An objection that can be made against *D3* is that it disregards the possibility that an individual that did not learn (one of) the languages of assessment during childhood might still have done so later in life. To take that objection into account we will use a slightly different definition in the following, namely:

D4: The respondent is classified as immigrant if *D3* applies **and** the language most often spoken at home by the respondent is not equal to (one of) the language(s) of assessment.

Both of the definitions *D3* and *D4* give rise to smaller immigrant shares than does the definition based on country of birth, *D2*.[18] However, the definition *D4* is more narrow than *D3* and, thus, results in the smallest immigrant shares in the population. On average, the immigrant share under *D4* is about half of that under *D3*. More specifically, under *D3*, ten of the 21 countries have immigrant shares in the population exceeding 10

[18] There is one single exception to this finding: for the Slovak Republic the language-based immigrant definitions *D3* and *D4* both yield larger immigrant population shares than does the definition based on country of birth, *D2*.

percent. In contrast, when the definition *D4* is applied the immigrant population shares exceed 10 percent in only three of the 21 countries; cf. Table 4. It can also be noted that under *D4* the immigrant share in one of the countries – Japan – is equal to zero.[19]

The definitions *D3* and *D4* also yield different immigrant mean scores. While the definition *D3* generates mean immigrant scores that are relatively close to those obtained under the definition *D2*, the definition *D4* results in mean scores that are distinctly lower than the corresponding scores under *D3* and, hence, under *D2.* In particular, whereas under *D3*, two countries have mean immigrant scores corresponding to proficiency level 1 (USA and France), this number increases to seven under *D4* (Sweden, USA, Norway, UK, Spain, Italy, and France), see Table 4. The higher number under *D4* is to be expected as individuals whose mostly used language coincides with (one of) the country's language(s) of assessment can be included among the immigrants under the definition *D3*, but not under *D4*. Everything else equal, individuals who master the language of assessment should be more likely to score high than individuals lacking this language competence. The spread in mean scores across countries is also larger for definition *D4* than for the definitions *D3* and *D2.*

It can be deduced from Table 4 that the differences in mean scores between immigrants and non-immigrants generally are negative under *D4*; the only exceptions are the Czech Republic and Poland.[20] This pattern is similar to the pattern for definition *D2* (and *D3*). However, the number of countries with large differences – exceeding one skill level, or 50 points – is larger in Table 4 (6, namely Sweden, Austria, Norway, Denmark, the Netherlands and France) than in Table 2 (2, Sweden and France).

Compared to Table 3, an interesting finding in Table 4 is that the differences in years of schooling between immigrants and non-immigrants are quite substantial in many countries. Specifically, the difference exceeds one year for seven countries; the corresponding number in Table 3 is two. Nevertheless, the coefficient of correlation for the differences between immigrants and non-immigrants in average scores and years of schooling under D4 is 0.69, i.e. very close to the corresponding correlation under *D2*, 0.67.

Comparing Table 3 and Table 4 it is not obvious which of the underlying immigrant definitions that will be associated with the largest (positive) differences in mean scores

[19] Japan's immigrant share is equal to 0.2 under *D2*, as well as under *D3*.
[20] These differences can be computed from Table 4 and they can also be obtained directly from Table 6.

among non-immigrants compared to the averages for the countries' entire populations. In general, the largest differences will be found for countries that both have large shares of immigrants and low mean scores among the immigrants. However, going from definition *D2* to definition *D4* typically means decreasing the immigrant shares *and* the immigrant mean score, changes that work in opposite directions.

Table A2 in the Appendix complements Table 4 by providing the corresponding mean scores for immigrants and non-immigrants with respect to literacy skills. The numbers in Table A2 are very similar to those in Table 4. For example, just like in Table 4, the largest skill differences between immigrants and non-immigrants are in Table A2 found for Sweden and France.

Table 4: Immigrant and non-immigrant characteristics, and mean numeracy scores when immigrants are defined according to definition D4.[1] Countries ordered in descending order by immigrant population shares

Country[2]	Immigrant population share, %	Immigrant mean numeracy score[3]	Non-immigrant mean numeracy score[3]	Immigrant mean years of schooling	Non-immigrant mean years of schooling
Canada	13.3	240.4	269.1	13.5	13.1
Sweden	10.6	216.9	286.6	11.4	12.1
USA	10.1	211.8	257.6	11.7	13.2
Austria	9.5	228.5	279.9	10.8	11.9
Norway	8.6	225.6	283.4	14.0	13.9
Ireland	6.8	237.8	256.9	15.2	14.4
Germany	6.2	226.2	274.8	11.3	12.6
Denmark	6.1	230.4	281.5	12.3	12.5
UK (England & Northern Ireland)	5.2	215.8	264.3	12.9	12.3
Netherlands	5.1	227.6	283.6	12.2	13.2
Spain	4.9	222.8	247.5	10.8	11.3
Italy	4.7	222.1	248.4	9.7	10.5
Belgium (Flanders)	4.6	245.5	282.6	11.9	12.4
Slovak Republic	4.2	244.0	277.2	11.7	13
France	4.0	188.3	257.0	8.9	11.3
Estonia	1.7	273.1	273.3	12.0	12.1
Finland	1.5	262.3	285.6	12.2	12.3
Czech Republic	0.8	282.4	275.7	14.6	12.9
Korea	0.5	235.9	263.6	12.1	12.6
Poland	0.1	297.2	259.8	13.9	12.5
Japan	0.0	–	288.2	–	12.9

[1] *D4*: The respondent is classified as immigrant if the first (or second) language that (s)he learned at home during childhood is not equal to (one of) the language(s) of assessment *and* the language most often spoken at home by the respondent is not equal to (one of) the languages of assessment.
[2] The countries Australia, Cyprus and the Russian Federation are not included in table. Access to the Australian data requires an approval from the Australian Bureau of Statistics, which we, currently, do not have. The Cypriot data have been withdrawn by the OECD. The data from the Russian Federation are preliminary and do not include the population in the Moscow area.
[3] For reference, the mean numeracy score in PIAAC equals 269 and the proficiency levels for numeracy are defined according to: 176 £ Level 1 < 226, 226 £ Level 2 < 276, 276 £ Level 3 < 326, 326 £ Level 4 < 376, and 376 £ Level 5.

Source: Own computations.

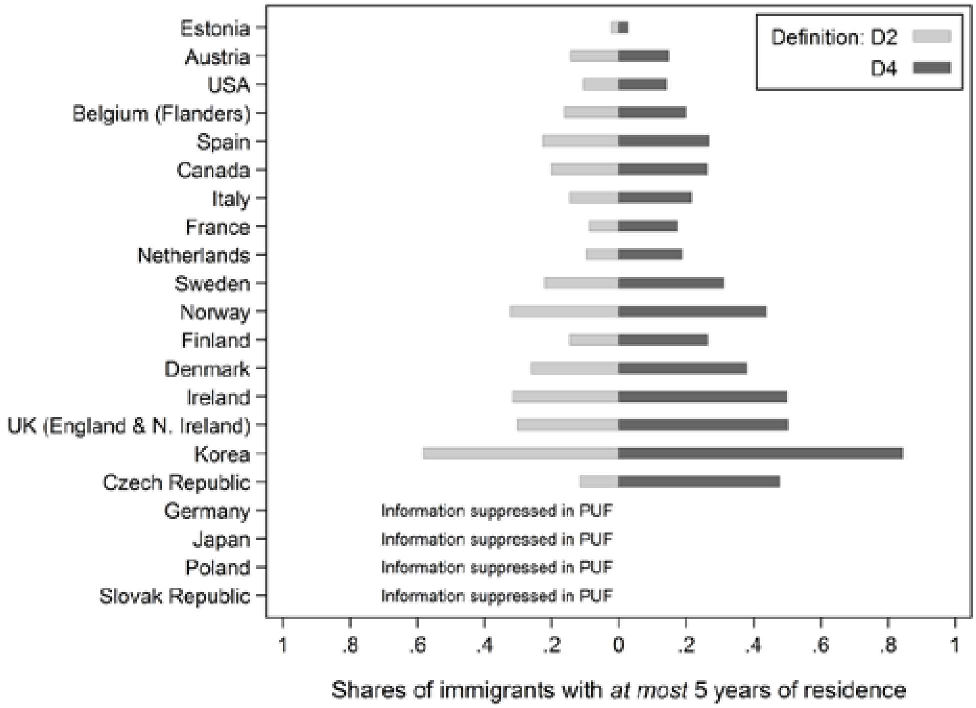
3.3 Is length of residence or age at arrival in the host country the key?

It might be that the differences between the immigrant definitions based on country of birth and spoken language, respectively, are due to some common underlying factor that is decisive with respect to both of the definitions. If so, how long the immigrant has been residing in the host country might be a candidate factor: length of residence will be crucial for the immigrant's skills in the host country language(s) as well as his/her knowledge about the country's culture. As argued by Blum et al. (2001), both may matter for the results in international skill surveys.

To investigate the importance of length of residence in the host country for our two immigrant definitions we consider two categories of immigrants: those whose length of residence in the host country is at most five years, and those whose length of residence is longer than five years.[21] Figure 1 shows the share of the former group, for both of the immigrant definitions and by country. It is clear from the figure that the share of immigrants with lengths of residence less than or equal to five years is larger when the language based immigrant definition is applied.

[21] The same partition is used in OECD (2013a, p. 126)

Figure 1:Shares of immigrants with length of residence in host country equal to at most 5 years, under the immigrant definitions D2 and D4. Countries ordered in ascending order according to the magnitude of the difference between the displayed immigrant shares under the definitions *D4* and *D2*, respectively



Note: For Germany, Japan, Poland, and the Slovak Republic the Public Use Files (PUFs) do not contain any information about immigrants' length of residence in the host country.

Source: Own computations

Now, if length of residence is a key factor behind the skill differences between non-immigrants and immigrants, then the fact that these skill differences in general are larger under the definition *D4* than under the definition *D2* should be associated with the fact that the share of immigrants that have lived at most five years in the host country is larger under *D4* than under *D2*. Specifically, for a given country, the difference:

$$\{m[S_{n\text{-}i,j}(D4)] - m[S_{i,j\text{-}}(D4)]\} - \{m[S_{n\text{-}i,j}(D2)] - m[S_{i,j\text{-}}(D2)]\}, \qquad (1)$$

where

$m[S_{n\text{-}i,j,}(Dk)]$ = the country's mean proficiency score among non-immigrants (n-i), indexed by j, under immigrant definition Dk, $k = 2,4$ (2)

and

$$m[S_{i,j-}(Dk)] = \text{the country's mean proficiency score among immigrants (i), indexed by j-, under immigrant definition } Dk,\ k = 2,4 \quad (3)$$

is expected to be positively related to the difference:

$$\text{share of } LoR_{i,j-}(D4) \leq 5 \text{ years}) - \text{share with } LoR_{i,j-}(D2) \leq 5 \text{ years} \quad (4)$$

where $LoR_{i,j-}$ denotes Length of Residence for immigrant j-. To check if this is indeed the case, we compute the correlation between these two differences, across the 17 countries for which there are data on the share of immigrants with at most five years of residence, cf. Figure 1. It turns out that the correlation is –0.39, and, thus, not positive as expected. However, it is only weakly significant at the 10 percent level. The corresponding correlation with respect to literacy is closer to zero, –0.20, and statistically insignficant.

Accordingly, length of residence cannot account for the differences in skills between non-immigrants and immigrants that our two immigrant definitions give rise to – at least not by itself. However, when we control for education and, thus, replace the above unconditional relation with the corresponding relation conditional on education, our results change quite a bit.

To control for education, we first regress the differences (1) and (4) on the following difference:

$$\{m[YoS_{n-i,j}(D4)] - m[YoS_{i,j-}(D4)]\} - \{m[YoS_{n-i,j}(D2)] - m[YoS_{i,j-}(D2)]\} \quad (5)$$

where YoS denotes Years of Schooling. The difference (5) can be computed by means of the information on years of schooling provided in Table 3 and Table 4. We then correlate the residuals from these two regressions. The resulting correlation becomes +0.42 with respect to numeracy skills, and +0.58 for literacy skills. Accordingly, the correlations have the expected sign and are also quite large.[22] Still, they show that even when we condition on education, length of residence cannot account for more than approximately half of the difference in skills between non-immigrants and immigrants. This indicates that the two immigrant definitions that we consider are truly distinct and captures different aspects on the relation between immigration and skills.

We have conducted the corresponding exercise replacing immigrants' length of residence with their age at arrival in the host country. Controlling for years of education we obtain only insignificant correlations leaving unchanged our conclusion that our two immigrant definitions are really distinct.

4 Immigration and countries' average skills

In this section we first show how a country's average score is related to the mean scores among its immigrants and non-immigrants, and how the difference between the country's average score and the average score of its non-immigrants can be computed. We then, in the next sub-section, present numbers showing how the country's estimated average skills changes when the immigrants are left out of the computations. We do this by country and for two of the immigrant definitions provided in Section 3: definitions *D2* and *D4*. Measures of statistical precision are provided along with the estimates of the average skill changes, which are also expressed in relative terms.

4.1 The computation of changes in average skills related to immigration

In analogy with (2) and (3) we define the mean proficiency score of a country's (entire) population according to:

$\mu[S_{p,J}(Dk)]$ = the mean proficiency score of the country's population, individuals indexed by J (j, $j \in J$), immigrant definition *Dk*. (6)

Further, we define

$\alpha(Dk)$ = the share of immigrants in the country's population under immigrant definition *k*. (7)

[22] The +0.42 correlation is significant at the 10 percent level of significance and the +0.58 correlation at the 1 percent level.

Using (2), (3), (6), and (7), we can express the mean score of the country's population as

$$\mathrm{m}[S_{p,J}(Dk)] = \mathrm{a}(Dk)\cdot \mathrm{m}[S_{i,j\text{-}}(Dk)] + [1 - \mathrm{a}(Dk)]\cdot \mathrm{m}[S_{n\text{-}i,j}(Dk)]. \qquad (8)$$

The change resulting when immigrants are left out of the computation of average proficiency can then be expressed according to:

$$\mathrm{m}[S_{n\text{-}i,j}(Dk)] - \mathrm{m}[S_{p,J}(Dk)] = \mathrm{a}(Dk)\cdot\{\mathrm{m}[S_{n\text{-}i,j}(Dk)] - \mathrm{m}[S_{i,j\text{-}}(Dk)]\}, \qquad (9)$$

i.e., the product of the immigrant population share and the difference in mean proficiency between non-immigrants and immigrants.

As a precursor to a statistical test associated with (9), notice that the total differential of $\mathrm{m}[S_{p,J}(Dk)]$ can be written:

$$\mathrm{d}\{\mathrm{m}[S_{p,J}(Dk)]\} = \mathrm{a}(Dk)\cdot \mathrm{d}\{\mathrm{m}[S_{i,j\text{-}}(Dk)]\} + [1 - \mathrm{a}(Dk)]\cdot \mathrm{d}\{\mathrm{m}[S_{n\text{-}i.j}(Dk)]\}$$
$$- \{\mathrm{m}[S_{n\text{-}i,j}(Dk)] - \mathrm{m}[S_{i,j\text{-}}(Dk)]\}\cdot \mathrm{d}[\mathrm{a}(Dk)]. \qquad (10)$$

Further, set

$$\mathrm{d}\{\mathrm{m}[S_{i,j\text{-}}(Dk)]\} = \mathrm{d}\{\mathrm{m}[S_{n\text{-}i,j}(Dk)]\} = 0 \qquad (11)$$

and

$$\mathrm{d}[\mathrm{a}(Dk)] = -\,\mathrm{a}(Dk), \qquad (12)$$

i.e., hold constant the mean of immigrants and non-immigrants and decrease the immigrant share to zero. This makes the right hand side of (10) exactly equal to the right hand side of (9), showing that under the constraints (11) – (12) the change $\mathrm{d}\{\mathrm{m}[S_{p,J}(Dk)]\}$ is equal to the difference $\mathrm{m}[S_{n\text{-}i,j}(Dk)] - \mathrm{m}[S_{p,J}(Dk)]$.

By the same argument, given the restrictions (11) – (12), testing the null hypothesis H_0: $\mathrm{d}\{\mathrm{m}[S_{p,J}(Dk)]\} = 0$ is equivalent to test H_0: $\mathrm{m}[S_{n\text{-}i,j}(Dk)] - \mathrm{m}[S_{p,J}(Dk)] = 0$. Further, these hypotheses are in turn equivalent to the null hypothesis

$$H_0: m[S_{n\text{-}i,j}(Dk)] = m[S_{i,j\text{-}}(Dk)] \quad (13)$$

because the equality in (13) is only way to make the expression

$$-\{m[S_{n\text{-}i,j}(Dk)] - m[S_{i,j\text{-}}(Dk)]\} \cdot d[a(Dk)]$$

zero, given (12).

Relying on asymptotic normality and exploiting the fact that the variables $S_{n\text{-}i,j}(Dk)$ and $S_{i,j\text{-}}(Dk)$ are stochastically independent, (13) can be tested by means of the statistic

$$z = \{m[S_{n\text{-}i,j}(Dk)] - m[S_{i,j\text{-}}(Dk)]\} / (\mathrm{Var}\{m[S_{n\text{-}i,j}(Dk)]\} + \mathrm{Var}\{m[S_{i,j\text{-}}(Dk)]\})^{1/2}. \quad (14)$$

Finally, to account for the fact that a given mean score point difference (9) is more important for a low-performing country than for a high-performing country we also compute the relative difference, i.e.,

$$(\{m[S_{n\text{-}i,j}(Dk)] - m[S_{p,J}(Dk)]\} / m[S_{p,J}(Dk)]) \times 100. \quad (15)$$

4.2 Changes in average skills by country and immigrant definitions

Tables 5 and 6 show the empirical counterparts of the formulae in the preceding subsection under the immigrant definitions *D2* and *D4*, respectively. It should be noted that in the tables the countries are ordered according to the numbers in the last column, the percentage differences between the average scores among non-immigrants and the entire population; cf. (15).

Both Table 5 and Table 6 show that while the differences in mean scores between non-immigrants and immigrants, $\{m[S_{n\text{-}i,j}(Dk)] - m[S_{i,j\text{-}}(Dk)]\}$, are very large for some countries – around 50 points, or one skill level – the differences in mean scores between non-immigrants and the entire population, $\{m[S_{n\text{-}i,j}(Dk)] - m[S_{p,J}(Dk)]\}$, albeit strongly significant, never exceed 10 points.[23] The relative differences, given (15), are even more modest. Only for Sweden, which exhibits the largest relative differences throughout,

[23] As noted in Section 4.1, testing whether the difference in mean scores among the non-immigrants and the population are statistically different from zero is, under the assumptions made, equivalent to testing to testing whether the difference in mean scores among non-immigrants and immigrants, respectively, are significant. Accordingly, the significance levels that apply to the former test apply to the latter test, as well.

does it exceed 2.5 percentage points under both definition *D2* and definition *D4*; cf. Table 5 and Table 6, respectively.

Table 5: Differences in numeracy skills between non-immigrants and immigrants according to definition D2 [1], and between non-immigrants and the entire population. Countries listed in descending order by relative skill differences between non-immigrants and the entire population

Country [2]	Skill difference between non-immigrants and immigrants	Immigrant population share	Skill difference between non-immigrants and population	Relative skill difference non-immgrants and population, %
Sweden	56.209*** [3]	0.175	9.837***	3.546
Norway	46.248***	0.134	6.197***	2.255
France	44.250***	0.128	5.664***	2.244
Netherlands	46.868***	0.128	5.999***	2.160
Canada	20.803***	0.257	5.346***	2.023
Germany	37.317***	0.138	5.150***	1.917
Austria	31.881***	0.163	5.197***	1.893
USA	31.373***	0.147	4.612***	1.850
Denmark	37.144***	0.118	4.383***	1.589
UK (England & N. Ireland)	27.372***	0.149	4.078***	1.584
Spain	21.747***	0.132	2.871***	1.187
Finland	25.769***	0.034	0.876***	1.096
Belgium (Flanders)	31.054***	0.069	2.143***	0.949
Estonia	15.361***	0.130	1.997***	0.783
Italy	17.160***	0.093	1.596***	0.648
Korea	32.999***	0.016	0.528***	0.204
Czech Republic	13.240*	0.040	0.530*	0.185
Slovak Republic	8.333*	0.023	0.192*	0.069
Poland	36.942***	0.002	0.074***	0.042
Japan	25.002	0.004	0.100	0.037
Ireland	-0.904	0.210	-0.190	-0.060

[1] *D2*: The respondent's country of birth is not equal to the country of assessment and there is non-missing information about the first (two) language(s) that (s)he learned at home during childhood and still understands, and the language that (s)he mostly speaks at home.
[2] The countries Australia, Cyprus and the Russian Federation are not included in table. Access to the Australian data requires an approval from the Australian Bureau of Statistics, which we, currently, do not have. The Cypriot data have been withdrawn by the OECD. The data from the Russian Federation are preliminary and do not include the population in the Moscow area.
[3] *, **, and *** represent statistical significance at the 10, 5, and 1 percent levels, respectively.

Source: Own computations.

Table 6: Differences in numeracy skills between non-immigrants and immigrants according to definition D4 [1], and between non-immigrants and the entire population. Countries listed in descending order by relative skill differences between non-immigrants and the entire population

Country [2]	Skill difference between non-immigrants and immigrants	Immigrant population share	Skill difference between non-immigrants and population	Relative skill difference non-immigrants and population, %
Sweden	69.652*** [3]	0.106	7.383***	2.691
USA	45.751***	0.101	4.621***	1.872
Norway	57.810***	0.086	4.972***	1.820
Austria	51.392***	0.095	4.882***	1.774
Canada	28.752***	0.133	3.824***	1.463
Finland	23.267***	0.015	0.349***	1.193
Netherlands	56.002***	0.051	2.856***	1.154
Denmark	51.025***	0.061	3.113***	1.142
Germany	48.626***	0.062	3.015***	1.127
France	68.725***	0.040	2.749***	1.101
UK (England & N. Ireland)	48.486***	0.052	2.521***	0.995
Belgium (Flanders)	37.112***	0.046	1.707***	0.803
Spain	24.706***	0.049	1.211***	0.684
Ireland	19.150***	0.068	1.302***	0.521
Slovak Republic	33.171***	0.042	1.393***	0.503
Italy	26.207***	0.047	1.232***	0.495
Estonia	0.258	0.017	0.004	0.077
Korea	27.645*	0.005	0.138*	0.068
Poland	-37.460	0.001	-0.037	-0.003
Czech Republic	-6.671	0.008	-0.053	-0.018
Japan	–	–	–	–

[1] *D4*: The respondent is classified as immigrant if the first (or second) language that (s)he learned at home during childhood is not equal to (one of) the language(s) of assessment, and the language most often spoken by the respondent at home is not equal to (one of) the language(s)s of assessment.
[2] The countries Australia, Cyprus and the Russian Federation are not included in table. Access to the Australian data requires an approval from the Australian Bureau of Statistics, which we, currently, do not have. The Cypriot data have been withdrawn by the OECD. The data from the Russian Federation are preliminary and do not include the population in the Moscow area.
[3] *, **, and *** represent statistical significance at the 10, 5, and 1 percent levels, respectively.

Source: Own computations.

Comparing the country-of-birth immigrant definition *D2* with the language-based definition *D4* we see that the latter in general yields larger differences in average skills between non-immigrants and immigrants, cf. the first columns in Table 5 and Table 6. However, this outcome is more than counteracted by the fact that under *D4* the countries immigrant shares in the population become much smaller than under *D2*.[24] The relative differences in the last columns of Table 5 and Table 6 are thus generally larger when definition *D2* is employed (Table 5) than when *D4* is used (Table 6). For most countries the differences are not very large, however. Notable exceptions are Ireland and the Slovak Republic where the relative differences under *D4* are *larger* than under *D2* and France and the Netherlands where they are distinctly smaller.

For literacy skills, the results corresponding to Table 5 and Table 6 are provided in Table A3 and Table A4, respectively, in Appendix 1. All of the comments made above apply to Table A3 and Table A4, too.

5 Immigration and country skill rankings

In Figure 2.6a in OECD (2013a) the PIAAC countries are ranked according to their mean numeracy proficiency scores, together with two indicators of the uncertainty associated with the ranking. The first indicator corresponds to assigning each of the countries to one of three categories, namely countries whose mean scores are i) significantly higher than the PIAAC average, ii) not significantly different from it, or iii) significantly lower than the PIAAC average, respectively. The second indicator lists, for each country, other countries whose mean scores are not significantly different from its own score.

The first of these indicators of statistical uncertainty is correct, the second is not. The reason why the second is in error is that it is based on comparisons of conventional 95 percent confidence intervals. For countries whose confidence intervals are not overlapping it is concluded that their mean scores are significantly different, else their scores are considered as not being statistically different. However, Goldstein and Healy (1995) show that conventional confidence intervals are inappropriate for such comparisons. Specifically, they demonstrate that the conventional confidence intervals can be either too wide – if the standard deviations for the mean scores of the different countries are very similar and few inter-country comparisons are conducted, or too narrow – if there are large differences across countries with respect to the standard deviations for the mean scores and if many inter-country comparisons are conducted.

[24] With one exception, the Slovak Republic, where the immigrant share increases from 2.3 percent under *D2* to 4.2 percent under *D4*.

To overcome this problem, we adopt a simulation procedure for the construction of confidence intervals, suggested by Leckie and Goldstein (2011), in the context of school league tables. Whether the standard deviations of the countries' average scores differ a lot or little does not matter for these simulation-based confidence intervals and they can be used for comparisons involving arbitrary numbers of pairs of countries. As long as the average country scores upon which the ranking is based can be treated as independent random variables, a comparison of the rank confidence intervals of any two countries will tell if their ranks are statistically significantly different (in which case the confidence intervals overlap) or not (in which case the confidence intervals do overlap). We describe the computation of the simulation-based confidence intervals in the first sub-section. In the second sub-section we extend the analysis by showing how the method can be applied to assess whether two skill rankings for one and the same country are statistically significantly different, namely its skill ranking with respect to the results of its non-immigrants and its entire population, respectively.

5.1 A simulation approach to rank uncertainty

To illustrate the method by Leckie and Goldstein (2011), we here consider the problem of estimating the uncertainty associated with the country ranking of average numeracy skills, in the entire population. We have estimates of means and variances for all countries. Assuming normality, these estimates provides us with estimates of the entire skill sampling distribution for each of the countries.

The sampling distributions can be viewed as the probability distributions for the ranges of numeracy skill scores that are possible for the respective countries.[25] All of the distributions will be normal but they will differ with respect to location (mean), range, and kurtosis (peakedness).

The simulation amounts to 10 000 iterations. For each iteration, we make random draws from all of the 21 country-specific sampling distributions and rank the countries based on these random draws. For every country the simulation thus yields 10 000 rank observations between 1 and 21 and a probability distributions on a subset of the [1,21] range. From these rank probability distributions we construct 95 percent confidence intervals by means of the 2.5th and 97.5th percentiles.

[25] That is to say, the ranges supported by the sampling distributions.

The result of the simulation is shown in Table 7, where for every given country a list of (other) countries is provided, in boldface, showing the countries whose ranks are not significantly different from the country in question. For comparison, a corresponding list, based on conventional (standard) 95 percent confidence intervals, where the names of the not significantly different countries have been underlined, is also provided.

Table 7: Statistical uncertainty in country ranks by average numeracy skills indicated by countries whose ranks are not significantly different at the 5 % level, according to Leckie and Goldstein (2011) – in boldface, and according to conventional confidence intervals – underlined

Rank[1]	Country[2]	Other countries whose ranks are not significantly different[3]
1	Japan	
2	Finland	Belgium (Flanders), Netherlands
3	Belgium (Flanders)	Denmark, Finland, Netherlands, Norway, Sweden
4	Netherlands	Belgium (Flanders), Denmark, Finland, Norway, Sweden
5	Sweden	Belgium (Flanders), Denmark, Netherlands, Norway
6	Norway	Austria, Belgium (Flanders), Czech Republic, Denmark, Netherlands, Slovak Republic, Sweden
7	Denmark	Belgium (Flanders), Netherlands, Norway, Sweden
8	Slovak Republic	Austria, Czech Republic, Norway
9	Czech Republic	Austria, Norway, Slovak Republic
10	Austria	Czech Republic, Estonia, Germany, Norway, Slovak Republic
11	Estonia	Austria, Germany
12	Germany	Austria, Estonia
13	Canada	Korea, UK (England & N. Ireland)
14	Korea	Canada, Poland, UK (England & N. Ireland)
15	UK (England & N. Ireland)	Canada, Korea, Poland
16	Poland	Ireland, Korea, UK (England & N. Ireland)
17	Ireland	France, Poland, USA
18	France	Ireland, USA
19	USA	France, Ireland, Italy
20	Italy	Spain, USA
21	Spain	Italy

[1] The rank is based on the country's average numeracy skills, in the entire population.
[2] The countries Australia, Cyprus and the Russian Federation are not included in table. Access to the Australian data requires an approval from the Australian Bureau of Statistics, which we, currently, do not have. The Cypriot data have been withdrawn by the OECD. The data from the Russian Federation are preliminary and do not include the population in the Moscow area.
[3] Countries whose ranks are not significantly different according to both the Leckie and Goldstein (2011) method and according to standard confidence intervals are written in boldface *and* underlined.

Source: OECD (2013a, Figure 2.6a) and own computations.

Table 7 shows that conventional confidence intervals *underestimate* the rank uncertainty. This follows from two results. First, the countries listed in the last column of

Table 7 are all written in boldface, which indicates that the uncertainty estimated by means of the standard confidence intervals is never larger than the uncertainty estimated by the Leckie and Goldstein (2011) method – had that been the case, the names of some countries would have been underlined but not written in boldface. Second, for 14 of the countries in the second column at least one of the countries listed in the third column is written in boldface but not underlined, indicating that the uncertainty estimated by the Leckie and Goldstein (2011) method is larger than the uncertainty estimated by means of the standard confidence intervals.[26]

Why is it that the conventional confidence intervals are too narrow and, thus, underestimate the rank uncertainties? As noted above, that will happen if there are large differences across countries with respect to the standard deviations for the mean scores and if many inter-country comparisons are conducted. Both of these conditions are satisfied. The standard deviations for the mean scores range from 0.53 for Estonia to 1.17 for the USA. And for 13 of the 21 countries the numbers of inter-country comparisons are three or higher.[27]

5.2 Confidence intervals for country ranks with and without immigrants

In this section, we want to test, for each country, if its rank with respect to its non-immigrants is statistically significantly different from the rank of its population. From equation (8) it can be seen that a necessary condition for the rank of the country's population to be significantly different from its rank with respect to its non-immigrants is that the ranks with respect non-immigrants and immigrants are significantly different, i.e. that the corresponding confidence intervals do not overlap. This is so because (8) tells that if this condition is not satisfied, the rank of the country's population will, in a statistical sense, be equal to its rank with respect to non-immigrants.

The necessary condition is, however, not sufficient. This can also be deduced by means of equation (8). Although (8) is formulated in terms of mean scores, its basic structure, defining the population as a weighted average of immigrants and non-immigrants, applies to any outcome measure. Accordingly, it applies to the outcome measure 'rank', in particular. Thus, if the ranks of non-immigrants and immigrants are

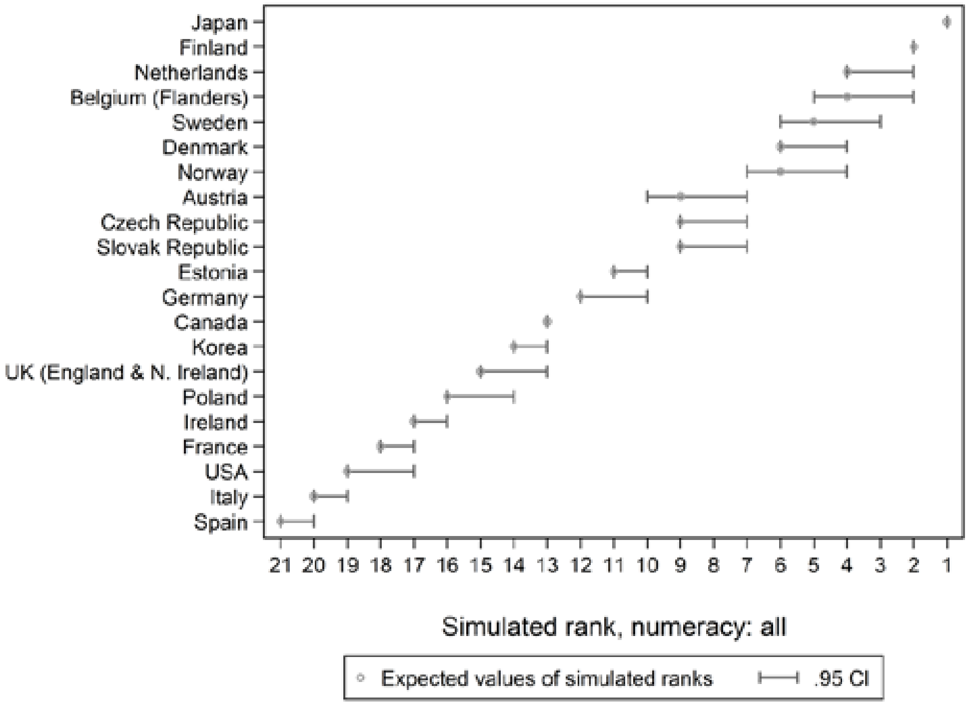
[26] The 14 countries are: the Netherlands, Norway, Denmark, the Slovak Republic, the Czech Republic, Austria, Germany, Canada, Korea, the UK (England & N. Ireland), Poland, Ireland, the USA, and Italy.
[27] The 13 countries are Belgium (Flanders), the Netherlands, Sweden, Norway, Denmark, the Slovak Republic, the Czech Republic, Austria, Korea, the UK (England & N. Ireland), Poland, Ireland, and the USA.

significantly different then the rank of the population must be given by the population-weighted average of the two ranks. Since the immigrant share α is well below 0.5, the weighted average rank and its associated confidence interval will be (much) closer to the non-immigrant rank confidence interval than to the immigrant rank confidence interval. And, so, it might well happen that the confidence interval for the weighted average rank overlaps with the confidence interval for the non-immigrant rank.

Going over to the empirics, we first provide simulation-based rank confidence intervals with respect to average scores in numeracy, for the countries' entire populations. We then compute the corresponding confidence intervals when immigrants are left out from the computation of the average scores. Finally, for each country we check if its two confidence intervals in the two rankings overlap. We do this for both of the two definitions of immigrants that we consider, i.e., *D2* and *D4*.

Figure 2 shows the expected values of the simulated ranks and 95 percent rank confidence intervals for numeracy skills in the countries' entire populations.

Figure 2: Expected values of simulated ranks [1] and 95 % rank confidence intervals for numeracy skills in the countries' [2] entire populations. Countries ordered in descending order according to average values of the simulated ranks



[1] For a given country, the rank observations are the ranks corresponding to the 10 000 draws from its' numeracy skill distribution for its entire population. The average simulated rank is given by the weighted mean of its' rank observations, the weights being the relative frequencies with which the different ranks have been observed. Both the averages and the lower and upper limits of the confidence intervals have been rounded to the nearest integer. For many countries, the rank distributions are very skewed, resulting in the average ranks coinciding with the upper or lower limits of the confidence intervals.

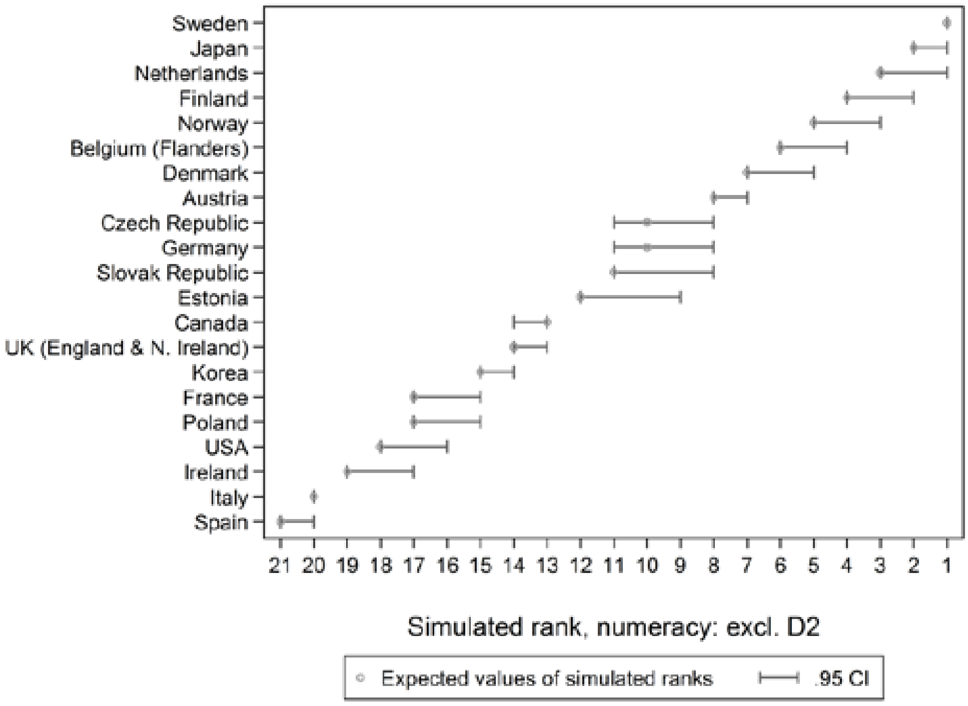
[2] The countries Australia, Cyprus and the Russian Federation are not included in table. Access to the Australian data requires an approval from the Australian Bureau of Statistics, which we, currently, do not have. The Cypriot data have been withdrawn by the OECD. The data from the Russian Federation are preliminary and do not include the population in the Moscow area.

Source: Own computations.

Figure 2 corresponds to Table 7. Accordingly, the counterparts to the overlaps of the confidence intervals in Figure 2 are the lists of countries in the last column of Table 7. For example, the expected value of the simulated ranks for the Netherlands is 4 and this rank is not significantly different from the ranks of Finland, Belgium (Flanders), Sweden, Norway, and Denmark.

Figure 3 shows the expected values of the simulated ranks and the 95 percent rank confidence intervals for numeracy skills when immigrants defined by *D2* are left out. It can immediately be seen that the ordering in Figure 3 is different from the ordering in Figure 2. For instance, the top-three countries in Figure 2 are Japan, Finland, and Belgium (Flanders), while the top-three in Figure 3 are Sweden, Japan, and the Netherlands. It can also be seen, however, that the rank uncertainties associated with the top-three in Figure 3 differ from the corresponding uncertainties in Figure 2. In general, for a given country both the expected value of its' simulated ranks and its' rank confidence interval differ between Figure 2 and Figure 3.

Figure 3: Expected values of simulated ranks [1] and 95 % rank confidence intervals for numeracy skills among non-immigrants, when immigrants are defined by definition D2. Countries [2] ordered in descending order according to average values of the simulated ranks



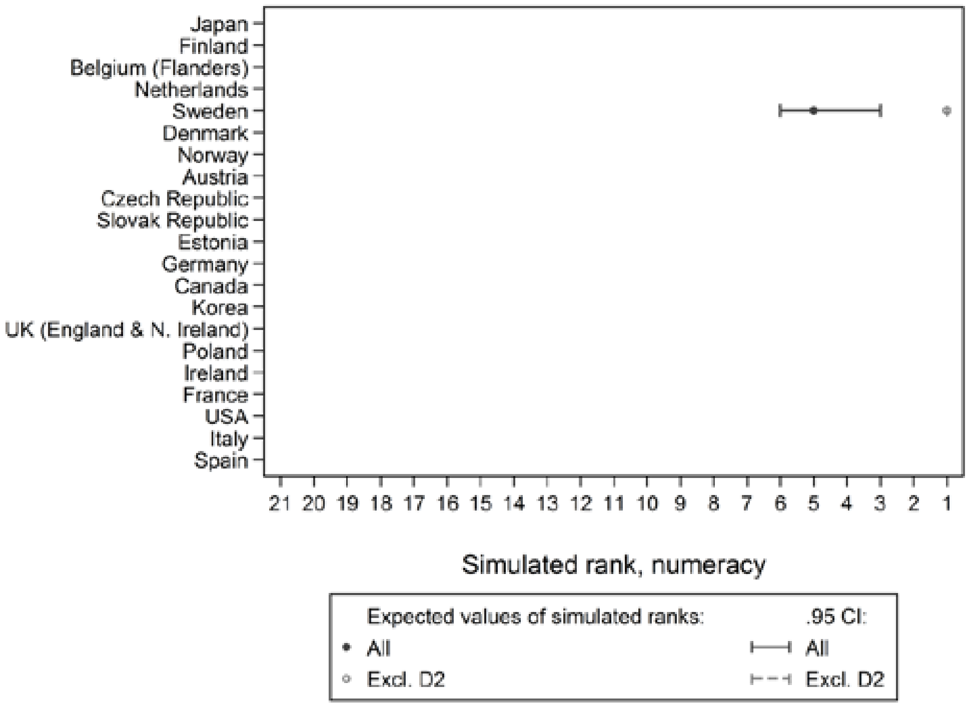
[1] For a given country, the rank observations are the ranks corresponding to the 10 000 draws from its numeracy skill distribution for its non-immigrants. The average simulated rank is given by the weighted mean of its rank observations, the weights being the relative frequencies with which the different ranks have been observed. Both the average and the lower and upper limits of the confidence intervals have been rounded to the nearest integer. For many countries, the rank distributions are very skewed, resulting in the average ranks coinciding with the upper or lower limits of the confidence intervals.
[2] The countries Australia, Cyprus and the Russian Federation are not included in table. Access to the Australian data requires an approval from the Australian Bureau of Statistics, which we, currently, do not have. The Cypriot data have been withdrawn by the OECD. The data from the Russian Federation are preliminary and do not include the population in the Moscow area.

Source: Own computations.

To judge whether leaving out the immigrants significantly changes a country's rank we compare the country's confidence intervals in Figure 2 and Figure 3 and check whether they overlap. If they do, we conclude that there is no significant change in the country's rank. If they do not, we infer that leaving out the country's immigrants significantly alters its rank. Figure 4 reports the result of such pairwise comparisons for all of the 21 countries. In the figure only significant results are reported, i.e. the non-overlapping confidence intervals.

Figure 4 shows that when immigrants are defined according to *D2* leaving them out in the computation of the countries' average numeracy skills barely affects the country skill ranking. Only for one country does the rank change significantly, namely for Sweden whose 95 percent rank confidence interval changes from [3,6] to [1].

Figure 4: Non-overlapping 95 % rank confidence intervals for numeracy skills, by country, from Figure 2 (solid line intervals) and Figure 3 (dashed line intervals), showing significant changes in country ranks when immigrants, defined by D2, are left out. Countries [1] ordered as in Figure 2, in descending order according to the simulated mean ranks in the entire population



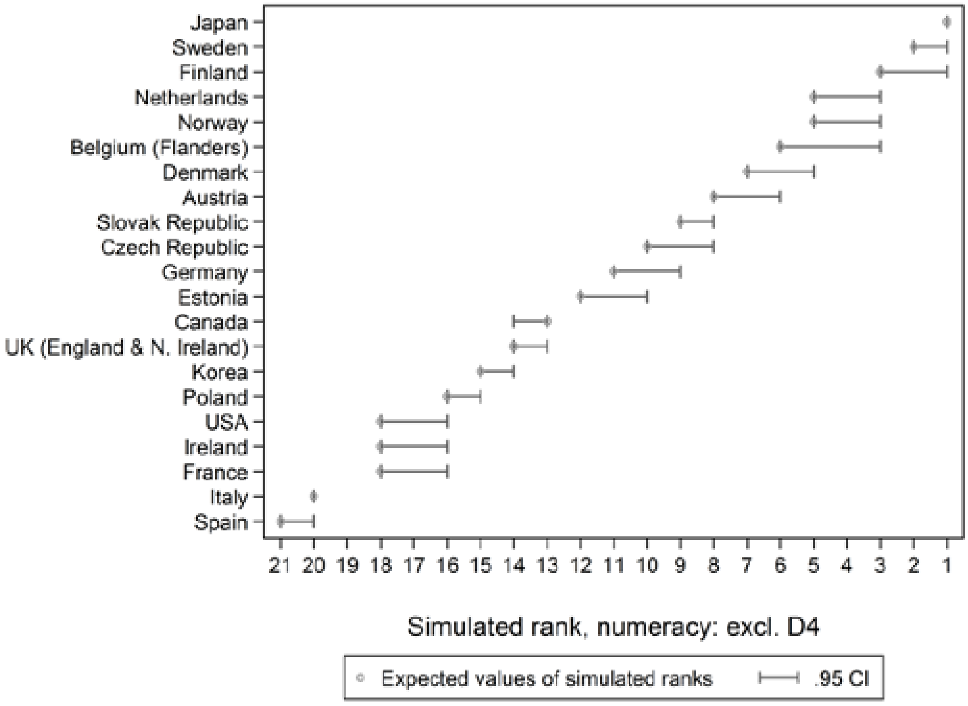
[1] The countries Australia, Cyprus and the Russian Federation are not included in table. Access to the Australian data requires an approval from the Australian Bureau of Statistics, which we, currently, do not have. The Cypriot data have been withdrawn by the OECD. The data from the Russian Federation are preliminary and do not include the population in the Moscow area.

Source: Own computations.

To see whether the result in Figure 4 is sensitive to the choice of definition of immigrants we repeat the exercise using immigrant definition *D4*. Figure 5 shows the simulated mean ranks and 95 percent rank confidence intervals for numeracy skills when immigrants defined by *D4* are left out.

Figure 5 looks similar to Figure 3, but there are some differences. For example, the ordering of the countries, based on mean simulated ranks, differs slightly with respect to the five highest ranking countries. And for some countries the width of the confidence intervals has increased (see, e.g., Sweden, Belgium (Flanders), and Austria) while for others it has decreased (for, e.g., Japan, the Slovak Republic, and the Czech Republic).

Figure 5: Expected values of simulated ranks [1] and 95 % rank confidence intervals for numeracy skills among non-immigrants, when immigrants are defined by definition D4. Countries [2] ordered in descending order according to average values of the simulated ranks



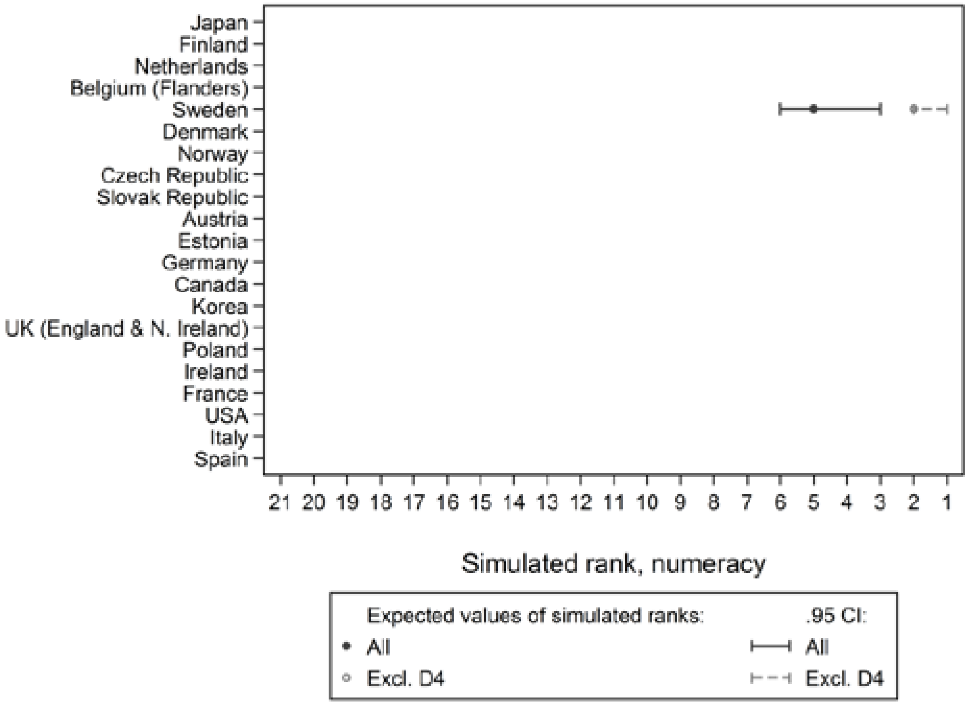
[1] For a given country, the rank observations are the ranks corresponding to the 10 000 draws from its numeracy skill distribution for its non-immigrants. The average simulated rank is given by the weighted mean of its' rank observations, the weights being the relative frequencies with which the different ranks have been observed. Both the average and the lower and upper limits of the confidence intervals have been rounded to the nearest integer. For many countries, the rank distributions are very skewed, resulting in the average ranks coinciding with the upper or lower limits of the confidence intervals.

[2] The countries Australia, Cyprus and the Russian Federation are not included in table. Access to the Australian data requires an approval from the Australian Bureau of Statistics, which we, currently, do not have. The Cypriot data have been withdrawn by the OECD. The data from the Russian Federation are preliminary and do not include the population in the Moscow area.

Source: Own computations.

Figure 6 shows that when the language-based immigrant definition *D4* is employed Sweden is, again, the only country for which the rank confidence intervals for non-immigrants and the entire population are not overlapping. For the rank of its entire population with respect to numeracy skills the 95 percent confidence interval is [3,6] while for its non-immigrants (implicitly) defined by *D4* the confidence interval is [1,2].

Figure 6: Non-overlapping 95 % rank confidence intervals for numeracy skills, by country, from Figure 2 (solid line intervals) and Figure 5 (dashed line intervals), showing significant changes in country ranks when immigrants, defined by D4, are left out. Countries [1] ordered as in Figure 2, in descending order according to the simulated mean ranks in the entire population



[1] The countries Australia, Cyprus and the Russian Federation are not included in table. Access to the Australian data requires an approval from the Australian Bureau of Statistics, which we, currently, do not have. The Cypriot data have been withdrawn by the OECD. The data from the Russian Federation are preliminary and do not include the population in the Moscow area.

Source: Own computations.

In summary, the conclusion is that the country ranking with respect to average numeracy skills is almost unaffected if immigrants are left out. And this inference is not sensitive to the choice of immigrant definition. Only for one country, Sweden, are significant changes – improvements – in rank recorded, for both of the immigrant definitions employed, i.e. *D2*, based on country of birth, and *D4*, based on whether language spoken coincides with (one of) the country’s assessment language.

We have repeated the exercise in this sub-section for literacy skills. This allows us to generalize our main conclusion, as follows: country skill rankings are almost unaffected if immigrants are left out. This inference is not sensitive either to choice of immigrant definition or type of skill considered.

Regarding the (very few) countries whose skill rankings are indeed affected we can add: the significant rank changes recorded for Sweden with respect to numeracy skills have no counterparts with respect to literacy skills. In fact, when literacy skills are considered only one significant rank change is found: under immigrant definition *D2*, France marginally improves it skill rank from [19] to [17,18], cf. Figure A1 in Appendix 1. Under immigrant definition *D4* there is no county at all for which the rank is improved when non-immigrants only are considered, as opposed to the entire population, cf. Figure A2.

6 Sensitivity analysis regarding language-related non-response

So far we have shown that our results are not sensitive to whether we define immigrants in terms of country of origin or use a language-based definition. The latter was used to delineate *non-immigrants* as individuals mastering (any of) the country's test assessment language(s). However, there is also another aspect on language that could be of importance and which is separate from the definitions of immigrants, namely the treatment of language-related non-response. As discussed in Section 2.2, countries that to a large extent handled language-related non-responding individuals by imputing test scores to them most likely obtained lower average scores by doing so, compared to if the individuals had not been assigned scores and, accordingly, had been left out of the computation of the country's average scores. This consideration is especially relevant for the only country whose ranking with respect to numeracy skills has been shown above to be significantly affected by immigration, i.e. Sweden. Sweden was one of only three countries in PIAAC that consistently handled language-related non-response by assignment of imputed scores, cf. Table 2.[28]

In this section we consider how the results for Sweden would be altered if all the individuals in the Swedish sample that were assigned imputed scores due to language-related non-response had instead had been left out of the computation of average scores. As discussed in Section 2.2, this would improve Sweden's average score, since the imputed scores generally are very low. Moreover, it would reduce the difference between the average scores for the Swedish non-immigrants and Sweden's full sample, as immigrants are over-represented among language-related non-respondents.

Altogether, 110 individuals in the Swedish sample were assigned imputed scores due to language-related non-response. While making up 2.5 percent of the Swedish

[28] The other two countries were Finland and Poland.

sample,[29] they only represent 1.3 percent of Sweden's population. That is to say, in the weighted sample their share is about half of what it is in the unweighted sample.

In terms of the immigrant definitions that we employ, the individuals with language-related imputed scores can be characterized as follows. All of them are foreign-born, i.e. immigrants according to definition *D2*. Almost 90 percent are classified as immigrants under definition *D4*. Thus, under the latter definition, 10 percent of the individuals with language-related imputed scores are classified as non-immigrants.

In Table 8 it can be seen that leaving out the respondents with language-related imputed scores decreases the share of immigrants in the population by about one percentage point, under both of the immigrant definitions considered. Moreover, the differences in average scores between non-immigrants and immigrants are substantially decreased, too; by around 9 score points under immigrant definition *D2* and about 12 points under definition *D4*. Both the decrease in the immigrant and the reduced non-immigrant vs immigrant score difference act to reduce the difference between the average scores of non-immigrants relative to the entire population. Nevertheless, the resulting changes are rather small: just above and just below 2 points, under definitions *D2* and *D4*, respectively or, in relative terms, 0.8 and 0.7 percentage points, cf. the last column of Table 8.

[29] 110 / 4469 = 0.0246

Table 8: For Sweden: differences in numeracy skills between non-immigrants and immigrants, and between non-immigrants and the entire population, when respondents with imputed scores due to language difficulties [1] are included and excluded, respectively. Reported according to *M* [*L*,*U*] where *M* is the simulated mean and *L* and *U* are the lower and upper bounds of the simulated 95 % confidence interval [2]

Sweden	Skill difference between non-immigrants and Immigrants	Immigrant population share	Skill difference between non-immigrants and population	Relative skill difference non-Immigrants and population, %
Definition D2				
Including imputed scores [3]	56.19 [51.90–60.39]	0.175	9.839 [9.099–10.59]	3.526
Excluding imputed scores	46.94 [42.60–51.32]	0.164	7.699 [6.969–8.428]	2.738
Definition D4				
Including imputed scores [4]	69.69 [63.53–75.89]	0.106	7.389 [6.731–8.058]	2.648
Excluding imputed scores	57.68 [51.23–64.43]	0.096	5.536 [4.897–6.174]	1.970

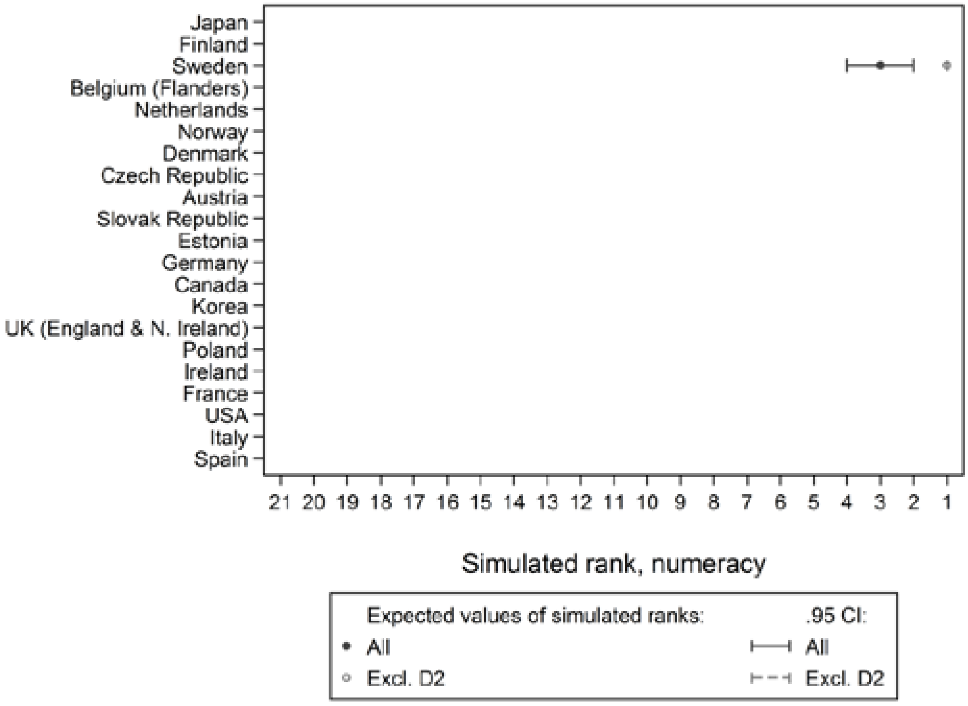
[1] Respondents with imputed scores due to language difficulties make up 2.5 % of the Swedish sample and represent 1.3 % of Sweden's population.
[2] For a given definition of immigrants, the simulated means and the lower and upper bounds of the simulated confidence intervals are based 10 000 draws from the assumed normal distribution of the differences for Sweden, when imputed scores are included and excluded, respectively.
[3] The small differences compared to the corresponding entries in Table 5 are due to the fact that the entries in this table are based on 10 000 simulations while the numbers in Table 5 are based on the (actual) values reported.
[4]. The small differences compared to the corresponding entries in Table 6 are due to the fact that the entries in this table are based on 10 000 simulations while the numbers in Table 6 are based on the (actual) values reported.

Source: Own computations.

We next consider whether there are any significant impacts on Sweden's ranking. To this end, we first exclude the individuals with language-related imputed scores from the Swedish sample. The remaining sample we then use to conduct analyses in analogy with those underlying Figures 2–6, without changing the samples of the other countries. For brevity, we only report on the last steps in these computations, i.e. on the significant rank differences, cf. Figure 7. Two things can be noted about this figure.

First, the exclusion of individuals with imputed scores due to language-related reasons improves Sweden's ranking when the countries' entire populations are considered. This can be seen by comparing the ordering of the countries in Figure 7 with the ordering in Figure 2. In both of these figures the ordering is based on the simulated mean rank of the countries with respect to their numeracy skills in the entire population. The only difference is that in Figure 7 Swedish individuals with language-related imputed scores are excluded. This results in an improvement of Sweden's expected rank from 5 to 3.

Figure 7: *For Sweden*: 95 % rank confidence interval for the entire population, excluding respondents with language-related imputed scores (solid line interval) are compared to the corresponding confidence intervals further excluding immigrants, defined by D2 [1], (dashed line intervals). *For all countries but Sweden*: 95 % rank confidence intervals for numeracy skills for the entire population, from Figure 2 (solid line intervals) are compared to the corresponding confidence intervals for the population excluding immigrants, defined by *D*2, from Figure 3 (dashed line intervals). Countries ordered in descending order according to expected values of simulated ranks in the entire population [2]



[1] Respondents with imputed scores due to language difficulties make up 2.5 % of the Swedish sample and represent 1.3 % of Sweden's population.
[2] The countries Australia, Cyprus and the Russian Federation are not included in table. Access to the Australian data requires an approval from the Australian Bureau of Statistics, which we, currently, do not have. The Cypriot data have been withdrawn by the OECD. The data from the Russian Federation are preliminary and do not include the population in the Moscow area.

Source: Own computations.

Second, Figure 7 shows that under the immigrant definition *D2*, Sweden's numeracy score rank among non-immigrants is still significantly different from the corresponding rank for Sweden's entire population, when the Swedish respondents with language-related imputed scores are excluded. However, the rank change, from [2,4] to [1], is smaller than when language-related imputed scores are included, in which case it is from [3,6] to [1], cf. Figure 4.

Under the language-based immigrant definition *D4*, exclusion of the Swedish respondents with language-related imputed scores results in a complete absence of significant differences in ranks among non-immigrants as opposed to non-immigrants and immigrants taken together.[30] Thus, the significantly higher rank for Sweden's non-immigrants than for its entire population that is shown in Figure 6 vanishes once the respondents with language-related imputed scores are excluded from the Swedish sample.

When applied to literacy skills, this sensitivity analysis shows that, unlike with respect to numeracy skills, the results for literacy are *not* sensitive to the treatment of language-related non-response. In particular, the figures obtained when the individuals with language-related non-response are left out of the sample look exactly like Figure A1 and Figure A2. [31]

To sum up: For Sweden, the results on country rank differences among non-immigrants, on the one hand, and non-immigrants and immigrants alike, on the other hand, are partly sensitive to the treatment of language-related non-response. Specifically, when the treatment is to assign imputed scores to the corresponding individuals and keep them in the sample, Sweden's rank among non-immigrants is significantly higher than the rank for Sweden's entire population, when numeracy skills are considered. When the treatment is to leave the individuals with language-related non-response out of the sample the significant rank difference either decreases – when immigrants are defined according to definition *D2*, or vanishes altogether – when immigrants are defined according to definition *D4*. In contrast, the results for literacy skills are not sensitive to the treatment of language-related non-response.

It can be argued that this sensitivity analysis is partial and that it overestimates the induced Swedish rank changes. Specifically, a more 'fair' comparison would involve the exclusion of individuals with language-related imputed scores from the samples of *all* of the countries participating in PIAAC, and not only from the Swedish sample. We agree with this objection. However, individual-level information about language-related non-response and imputed scores is not publicly available.

[30] To save space, we do not show the corresponding figure.
[31] For brevity, these results are not documented in Appendix 1.

7 Concluding comments

Theory is of limited guidance when it comes to the relation between immigration and adult skills. Whether a country's human capital benefits or suffers from immigration depends on many immigrant-specific and host country specific factors, as well as time-specific conditions.

This paper suggests that simple numerical measures are likely to be of limited guidance, too. For example, the large differences in average literacy skills between immigrants and natives, to the natives' advantage, reported in OECD (2013a) for many of the countries participating in PIAAC, might be taken to be informative about the consequences of immigration for the skills of a country's entire population. However, our analysis makes it clear that these differences are not sufficient statistics for how countries' average scores are influenced by immigration. And their connection to the countries' relative positions in the international skill distribution is even more elusive.

Our analysis also shows that when it comes to assessing changes in country rankings it is very important to account for the statistical uncertainties involved. Moreover, we have demonstrated that standard confidence intervals underestimate the rank uncertainties, compared to an appropriate method for their computation that has been proposed by Leckie and Goldstein (2011).

For almost all countries participating in PIAAC, we find that average numeracy and literacy skills are statistically significantly higher when immigrants are excluded compared to when the country's entire population is considered. However, the differences are not very substantial. The largest difference, found for Sweden, is 10 score points, from 279 to 289, or 3.5 percent. These numbers are obtained when immigrants are defined as foreign-born individuals. When immigrants are defined as individuals not speaking (any of) the language(s) of assessment the differences become somewhat smaller. Again, however, the Swedish difference is the largest: 7 points or 2.7 percent.

Country rankings based on non-immigrants only are shown to be very similar to the country rankings based on the countries' entire populations. With respect to numeracy skills, only one country – Sweden – obtains a significantly better ranking when the immigrants are left out. This holds under both of the immigrant definitions considered. In each case Sweden's rank confidence interval is changed from rank [3,6] to a rank interval including only the top rank. However, a sensitivity analysis shows that the Swedish results are partly dependent upon the fact that Sweden, unlike almost all other countries, assigned imputed (low) scores to all individuals that did not take the skill assessments tests due to language difficulties. Under the alternative – frequently employed by most of the participating countries – of simply disregarding these individuals in the computation of average scores, Sweden's significant rank differences between non-immigrants and the entire population are either markedly reduced or vanish altogether, depending on the definition of immigrants employed.

With respect to literacy skills the country rankings based on non-immigrants only are even more similar to the corresponding rankings based on the countries' entire populations than for numeracy skills. France is the only country obtaining a better ranking when the countries' non-immigrants are considered, instead of their entire populations. Moreover, the improvement is marginal: the rank confidence interval changes from rank [19] to rank [17,18] and occurs only under the immigrant definition based on country of birth (*D2*). These results are not sensitive to the treatment of language-related non-response in Sweden.

A one-sentence conclusion from our results is the following: If you are not satisfied with your country's performance in PIAAC, don't blame the immigrants![32]

With respect to future waves of PIAAC, there are two lessons from this paper. First, when the results are presented country rank uncertainties should be computed according to the method suggested by Leckie and Goldstein, as in this paper. Given the paramount interest in country skill rankings and our finding that the standard confidence intervals employed in OECD (2013a) underestimate the uncertainty in the rankings, this would be an important improvement. Secondly, efforts should be taken to streamline the treatment of language-related non-response across countries. This would increase comparability across countries with respects to results as well as with respect to resources devoted to the survey.

[32] We owe this succinct statement to Sari Sulkunen.

References

Blum, A., H. Goldstein, and F. Guérin-Pace (2001), “International Adult Literacy Survey (IALS): An analysis of international comparisons of adult literacy”, *Assessment in Education: Principles, Policy & Practice,* 8(2), 225-246.

Bonnet, G. (2002), “Reflections in a Critical Eye: On the pitfalls of international assessment”, *Assessment in Education: Principles, Policy & Practice*, 9(3), 387-399.

Brunello, G. and L. Rocco (2013), “The effect of immigration on the school performance of natives: Cross country evidence using PISA test scores”, *Economics of Education Review*, 32, 234-246.

Entorf, H. and N. Minou (2005), “What a Difference Immigration Policy Makes: A comparison of PISA Scores in Europe and Traditional Countries of Immigration”, *German Economic Review*, 6(3), 355-376.

Dunn, O. J. (1961), “Multiple Comparisons Among Means”, *Journal of the American Statistical Association*, 56(293), 52-64.

Goldstein, H. and M. J. R. Healy (1995), “The Graphical Presentation of a Collection of Means”, *Journal of the Royal Statistical Society, Series A*, 158(1), 175-177.

Jensen, P. and A. W. Rasmussen (2011), “The effect of immigrant concentration in schools on native and immigrant children’s reading and math skills”, *Economics of Education Review*, 30, 1503-1515.

Kahn, L. M. (2004), “Immigration, skills and the labor market: International evidence”, *Journal of Population Economics*, 17, 501-534.

Korpi, T. (2012), “Importing Skills: Migration Policy, Generic Skills and Earnings among Immigrants in Australia, Europe and North America”. In J. Leschke et al. (Eds.) *Migration and Labour Markets in Troubled Times: Skills Mismatch, Return Migration and Policy Responses*, Ashgate: Surrey.

Leckie, G. and H. Goldstein (2011), “Understanding Uncertainty in School League Tables”, *Fiscal Studies*, 32(2), 207-224.

OECD (2000), *Literacy in the Information Age: Final Report of the International Adult Literacy Survey,* OECD Publishing: Paris.

OECD (2012), *Untapped Skills: Realizing the potential of immigrant students*, OECD Publishing: Paris.

OECD (2013a), *OECD Skills Outlook 2013, First Results from the Survey of Adult Skills*, OECD Publishing: Paris.

OECD (2013b), *The Survey of Adult Skills: Reader's Companion*, OECD Publishing: Paris.

OECD (2013c), *Technical Report of the Survey of Adult Skills (PIAAC)*, OECD Publishing: Paris.

OECD (2016), *Skills Matter: Further Results from the Survey of Adult Skills*, OECD Skill Studies, OECD Publishing: Paris.

Rangvid, B. S. (2010), "Source country differences in test score gaps: evidence from Denmark", *Education Economics*, 18(3), 269-295.

Sahlgren, G. H. (2015), "Invandringen och Sveriges resultatfall i Pisa" (Immigration and the Swedish decrease in average scores in Pisa), IFN Policy Paper Nr 71, Research Institute of Industrial Economics: Stockholm.

Skolverket (2016), "Invandringens betydelse för skolresultatet" (The importance of immigration for scholastic achievements in Sweden), Skolverket: Stockholm.

van der Linden, W. J. and R. K. Hambleton (Eds.) (1997), *Handbook of Modern Item Response Theory*, Springer Verlag: New York.

Appendix 1: Tables and figures relating to literacy skills

Table A1: Immigrant and non-immigrant characteristics, and mean literacy scores when immigrants are defined according to definition D2.[1] Countries ordered in descending order by immigrant population shares

Country [2]	**Immigrant population share, %**	**Immigrant mean literacy score [3]**	**Non-immigrant mean literacy score [3]**	**Immigrant mean years of schooling**	**Non-immigrant mean years of schooling**
Canada	25.7	255.6	279.5	13.8	12.9
Ireland	21.0	262.8	267.5	15.3	14.3
Sweden	17.5	235.0	288.7	11.7	12.1
Austria	16.3	247.9	273.7	11.7	11.8
UK (England & Northern Ireland)	14.9	255.1	275.6	13.0	12.3
USA	14.7	239.4	275.1	12.6	13.2
Germany	13.8	240.8	274.5	11.9	12.6
Norway	13.4	245.6	283.6	14.1	13.9
Spain	13.2	232.0	254.8	11.2	11.3
Estonia	13.0	256.2	279.0	12.0	12.1
France	12.8	229.6	266.9	10.1	13.2
Netherlands	12.8	246.9	289.5	12.6	11.4
Denmark	11.8	237.7	275.2	12.5	12.5
Italy	9.3	228.2	252.8	10.3	10.5
Belgium (Flanders)	6.9	245.4	278.3	12.0	12.4
Czech Republic	4.0	266.8	274.3	13.1	12.9
Finland	3.4	262.3	290.6	12.4	12.2
Slovak Republic	2.3	268.3	274.0	12.6	12.9
Korea	1.6	235.4	273.2	11.9	12.6
Japan	0.4	272.8	296.3	12.6	12.9
Poland	0.2	265.8	266.9	12.8	12.5

[1] *D2*: The respondent's country of birth is not equal to the country of assessment and there is non-missing information about the first (two) language(s) that (s)he learned at home during childhood and still understands, and the language that (s)he mostly speaks at home.

[2] The countries Australia, Cyprus and the Russian Federation are not included in table. Access to the Australian data requires an approval from the Australian Bureau of Statistics, which we, currently, do not have. The Cypriot data have been withdrawn by the OECD. The data from the Russian Federation are preliminary and do not include the population in the Moscow area.

[3] For reference, the mean numeracy score in PIAAC equals 269 and the proficiency levels for numeracy are defined according to: Below Level 1 < 176, 176 £ Level 1 < 226, 226 £ Level 2 < 276, 276 £ Level 3 < 326, 326 £ Level 4 < 376, and 376 £ Level 5.

Source: Own computations.

Table A2: Immigrant and non-immigrant characteristics, and mean literacy scores when immigrants are defined according to definition D4.[1] Countries ordered in descending order by immigrant population shares

Country[2]	Immigrant population share, %	Immigrant mean literacy score[3]	Non-immigrant mean literacy score[3]	Immigrant mean years of schooling	Non-immigrant mean years of schooling
Canada	13.3	244.7	277.8	13.5	13.1
Sweden	10.6	219.4	286.4	11.4	12.1
USA	10.1	226.3	274.8	11.7	13.2
Austria	9.5	230.2	273.6	10.8	11.9
Norway	8.6	234.8	282.6	14.0	13.9
Ireland	6.8	242.4	268.3	15.2	14.4
Germany	6.2	230.6	272.5	11.3	12.6
Denmark	6.1	222	274	12.3	12.5
UK (England & Northern Ireland)	5.2	231.6	274.8	12.9	12.3
Netherlands	5.1	236.2	286.9	12.2	13.2
Spain	4.9	225.7	253.6	10.8	11.3
Italy	4.7	217.3	252.1	9.7	10.5
Belgium (Flanders)	4.6	239.2	277.8	11.9	12.4
Slovak Republic	4.2	253	274.8	11.7	13
France	4.0	204.7	264.5	8.9	11.3
Estonia	1.7	272.6	276.1	12.0	12.1
Finland	1.5	260.3	290.9	12.2	12.3
Czech Republic	0.8	280.4	274	14.6	12.9
Korea	0.5	233.9	272.8	12.1	12.6
Poland	0.1	328.4	266.9	13.9	12.5
Japan	0.0	-	296.3	–	12.9

[1] *D4*: The respondent is classified as immigrant if the first (or second) language that (s)he learned at home during childhood is not equal to (one of) the language(s) of assessment *and* the language most often spoken at home by the respondent is not equal to (one of) the languages of assessment.

[2] The countries Australia, Cyprus and the Russian Federation are not included in table. Access to the Australian data requires an approval from the Australian Bureau of Statistics, which we, currently, do not have. The Cypriot data have been withdrawn by the OECD. The data from the Russian Federation are preliminary and do not include the population in the Moscow area.

[3] For reference, the mean numeracy score in PIAAC equals 269 and the proficiency levels for numeracy are defined according to: 176 £ Level 1 < 226, 226 £ Level 2 < 276, 276 £ Level 3 < 326, 326 £ Level 4 < 376, and 376 £ Level 5.

Source: Own computations.

Table A3: Differences in literacy skills between non-immigrants and immigrants according to definition D2 [1], and between non-immigrants and the entire population. Countries listed in descending order by relative skill differences between non-immigrants and the entire population

Country [2]	**Skill difference between non-immigrants and immigrants**	**Immigrant population share**	**Skill difference between non-immigrants and population**	**Relative skill difference non-immigrants and population, %**
Sweden	53.64*** [3]	0.175	9.387***	3.378
Canada	23.82***	0.257	6.122***	2.244
USA	35.64***	0.147	5.239***	1.957
Netherlands	42.58***	0.128	5.451***	1.936
Norway	38.02***	0.134	5.095***	1.852
France	37.30***	0.128	4.774***	1.811
Germany	33.70***	0.138	4.651***	1.746
Denmark	37.52***	0.118	4.427***	1.648
Austria	25.77***	0.163	4.200***	1.559
Spain	22.80***	0.132	3.009***	1.213
UK (England & N. Ireland)	20.49***	0.149	3.053***	1.144
Estonia	22.82***	0.130	2.966***	1.119
Finland	28.38***	0.034	0.965***	1.075
Belgium (Flanders)	32.86***	0.069	2.267***	1.025
Italy	24.55***	0.093	2.283***	0.910
Ireland	4.688**	0.210	0.984**	0.369
Korea	37.77***	0.016	0.604***	0.224
Czech Republic	7.439	0.040	0.298	0.096
Slovak Republic	5.656	0.023	0.130	0.047
Japan	23.49*	0.004	0.094*	0.034
Poland	1.154	0.002	0.002	0.013

[1] *D2*: The respondent's country of birth is not equal to the country of assessment and there is non-missing information about the first (two) language(s) that (s)he learned at home during childhood and still understands, and the language that (s)he mostly speaks at home.
[2] The countries Australia, Cyprus and the Russian Federation are not included in table. Access to the Australian data requires an approval from the Australian Bureau of Statistics, which we, currently, do not have. The Cypriot data have been withdrawn by the OECD. The data from the Russian Federation are preliminary and do not include the population in the Moscow area.
[3] *, **, and *** represent statistical significance at the 10, 5, and 1 percent levels, respectively.

Source: Own computations.

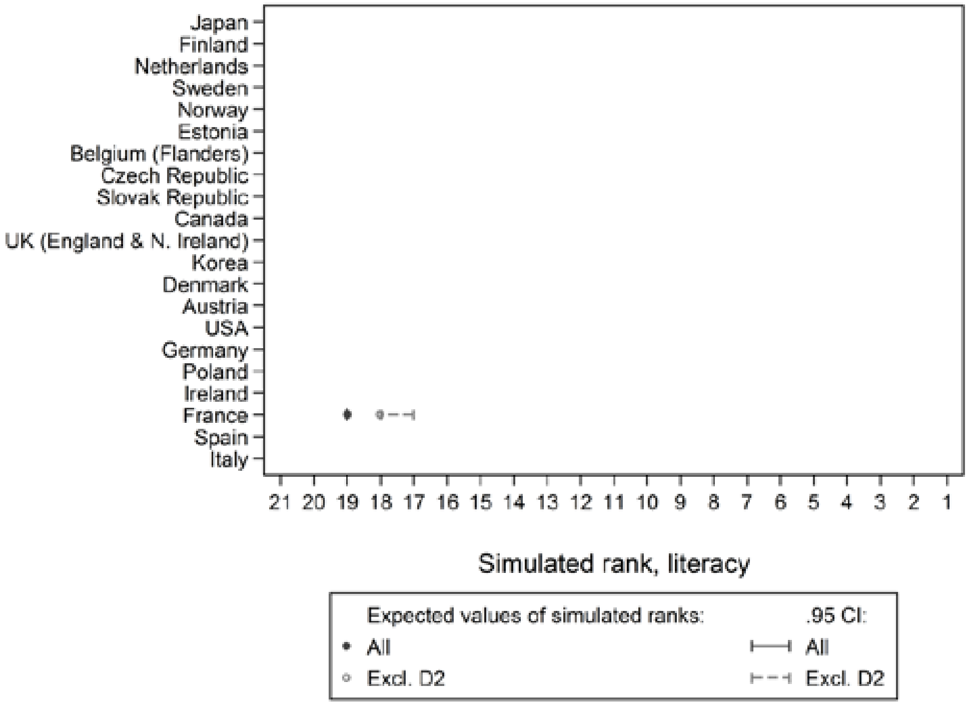
Table A4: Differences in literacy skills between non-immigrants and immigrants according to definition D4 [1], and between non-immigrants and the entire population. Countries listed in descending order by relative skill differences between non-immigrants and the entire population

Country [2]	**Skill difference between non-immigrants and immigrants**	**Immigrant population share**	**Skill difference between non-immigrants and population**	**Relative skill difference non-immigrants and population, %**
Sweden	66.99*** [3]	0.106	7.101***	2.577
USA	48.48***	0.101	4.896***	1.840
Canada	33.14***	0.133	4.408***	1.632
Austria	43.37***	0.095	4.120***	1.526
Norway	47.83***	0.086	4.113***	1.504
Denmark	52.04***	0.061	3.174***	1.195
Finland	30.51***	0.015	0.458***	1.153
Netherlands	50.69***	0.051	2.585***	1.022
Germany	41.83***	0.062	2.593***	0.980
France	59.77***	0.040	2.391***	0.904
Belgium (Flanders)	38.65***	0.046	1.778***	0.858
UK (England & N. Ireland)	43.18***	0.052	2.245***	0.853
Spain	27.90***	0.049	1.367***	0.725
Ireland	25.85***	0.068	1.758***	0.657
Italy	34.77***	0.047	1.634***	0.643
Slovak Republic	21.77***	0.042	0.914***	0.333
Estonia	3.473	0.017	0.059	0.081
Korea	38.89***	0.005	0.194***	0.079
Poland	-61.52	0.001	-0.062	-0.008
Czech Republic	-6.435	0.008	-0.051	-0.021
Japan	–	0.000	–	–

[1] *D4*: The respondent is classified as immigrant if the first (or second) language that (s)he learned at home during childhood is not equal to (one of) the language(s) of assessment, and the language most often spoken by the respondent at home is not equal to (one of) the language(s)s of assessment.
[2] The countries Australia, Cyprus and the Russian Federation are not included in table. Access to the Australian data requires an approval from the Australian Bureau of Statistics, which we, currently, do not have. The Cypriot data have been withdrawn by the OECD. The data from the Russian Federation are preliminary and do not include the population in the Moscow area.
[3] *, **, and *** represent statistical significance at the 10, 5, and 1 percent levels, respectively.

Source: Own computations.

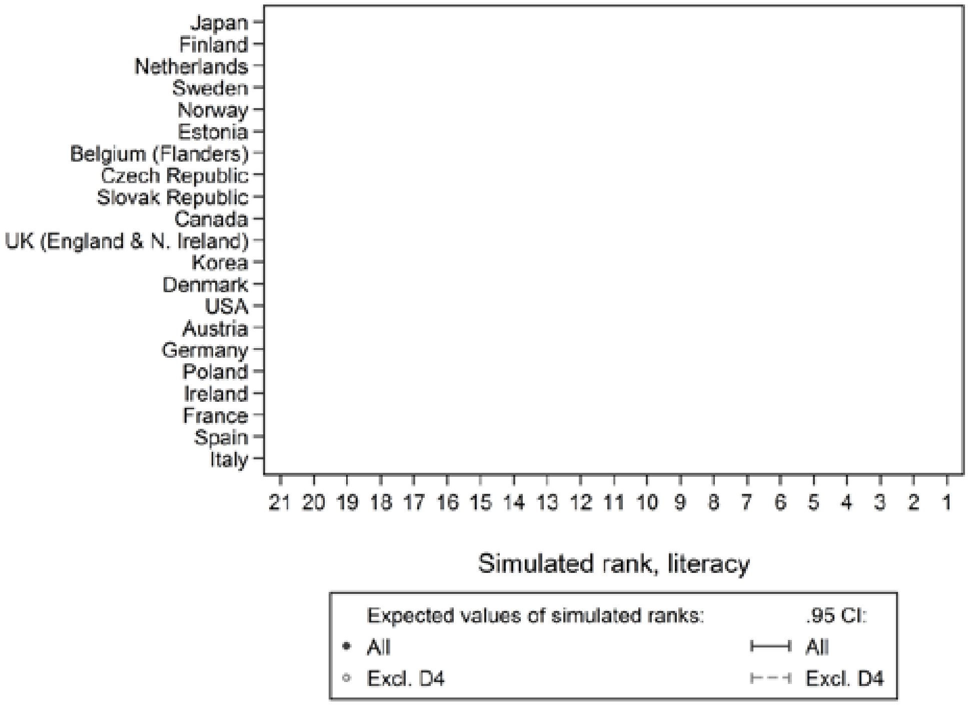
Figure A1: Non-overlapping 95 % rank confidence intervals for literacy skills, by country, for the entire population (solid line intervals) and for non-immigrants only, when immigrants are defined by D2 (dashed line intervals). Countries [1] ordered in descending order according to the simulated mean ranks in the entire population



[1] The countries Australia, Cyprus and the Russian Federation are not included in table. Access to the Australian data requires an approval from the Australian Bureau of Statistics, which we, currently, do not have. The Cypriot data have been withdrawn by the OECD. The data from the Russian Federation are preliminary and do not include the population in the Moscow area.

Source: Own computations.

Figure A2: Non-overlapping 95 % rank confidence intervals for literacy skills, by country, for the entire population (solid line intervals) and for non-immigrants only, when immigrants are defined by D4 (dashed line intervals). Countries [1] ordered in descending order according to the simulated mean ranks in the entire population



[1] The countries Australia, Cyprus and the Russian Federation are not included in table. Access to the Australian data requires an approval from the Australian Bureau of Statistics, which we, currently, do not have. The Cypriot data have been withdrawn by the OECD. The data from the Russian Federation are preliminary and do not include the population in the Moscow area.

Source: Own computations.

Appendix 2: Computation of variances allowing for both sampling error and skill measure uncertainty

PIAAC analyses involving plausible values of skill proficiency require 810 computations, in order to properly account for both sampling and measurement uncertainties. First, to account for measurement error, and get the correct point estimate of the statistic of interest, 10 computations using the final sampling values are needed, one per plausible value. Next, to account for the sampling uncertainty, and get correct standard errors, 10 × 80 computations have to be conducted, using a set of 80 jackknife replicate weights on each plausible value. [33]

The ten plausible values are random draws from each individual's proficiency distribution, where the distribution is estimated using the individual's own item responses as well as the responses of similar individuals. As noted in the main text, no respondent in PIAAC was tested on all skill domains and neither on all items within a domain. Methodologically, this approach is based on *Item Response Theory* (IRT), cf. van der Linden and Hambleton (1997).

The jackknife replicate weights create re-calculated subsamples, mirroring the design of the full sample, which are used to get an estimate of the variability of the results had the final sample been slightly different. The much more costly (and usually unfeasible) option would have been to sample from the population over and over again. Two types of jackknife methods were used in PIAAC: i) delete-one (delete-a-group or random groups approach) jackknife, known as JK1, and ii) paired jackknife, known as JK2. The choice of method depended on the sampling design of each country and followed from detailed guidelines on how to create the replicate weights. The majority of countries used JK2. (OECD, 2013c)

The formula for the computation of standard errors for a statistic of interest, ε, is:

$$SE_{\varepsilon} = \sqrt{\left[\sum_{p=1}^{P}\left(f \times \sum_{r=1}^{R}\left(\varepsilon_{r,p} - \varepsilon_{0,p}\right)^{2}\right) \times \frac{1}{P}\right] + \left[\left(1 + \frac{1}{P}\right) \times \frac{\sum_{p=1}^{P}\left(\varepsilon_{0,p} - \bar{\varepsilon}_{0,P}\right)^{2}}{P-1}\right]}$$

where

[33] Two countries have fewer sets of replicate weights resulting in fewer computation: the USA has 45 and the Russian Federation has 12 replicate weights.

$$\bar{\varepsilon}_{0,P} = \frac{\sum_{p=1}^{P} \varepsilon_{0,p}}{P}$$

and *p* and *r* denote plausible value and replicate weight, respectively. In PIAAC, the total number of plausible values, *P*, is always ten and the total number of replicates, *R*, is commonly equal to 80. Finally, *f* is a multiplier constant equal to one for JK2 countries while for JK1 countries $f = \frac{R-1}{R}$.