



Received: 28 May 2014
Accepted: 12 August 2014
Published: 29 September 2014

*Corresponding author: Kaycheng Soh, National Institute of Education, Nanyang Technological University, Singapore, Singapore
E-mail: sohkaycheng@hotmail.com

Reviewing editor:
Bob Adamson, Hong Kong Institute of Education, Hong Kong

Additional article information is available at the end of the article

INTERNATIONAL & COMPARATIVE EDUCATION | RESEARCH ARTICLE

Test language effect in international achievement comparisons: An example from PISA 2009

Kaycheng Soh^{1*}

Abstract: International achievement comparison studies assess students on core subjects such as Reading, Mathematics and Science. Students who do not speak the test language at home can be expected to be disadvantaged because of language proficiency. The test language effect has not been given sufficient attention. The present study investigated probable test language effect by using as data the country means reported in Reading, the PISA 2009 Reading. There was a wide range of proportions of non-speaker of test language among the participating countries. The average proportion of test language speakers is 80% with a wide standard deviation of 22%. The Reading mean for test language speakers is 39.2 points greater than that for non-speakers and the effect size is Cohen's $d = .69$. An adjusted Reading means to off-set test language effect was suggested. Careful scrutiny of the differences between original and adjusted means indicates that the test language effect is not simply linear. Effectiveness in second-language teaching may account for this complexity. Further research is indicated.

Subjects: General Language Reference, International & Comparative Education, Language Teaching & Learning

Keywords: PISA, test language, second language

1. Introduction

For large-scale international achievement comparison studies like Programme for International Student Assessment (PISA; OECD, 2010a), Trends in International Mathematics and Science Study (TIMSS; Mullis, Martin, Foy, & Arora, 2012) and Progress in International Reading Literacy Study (PIRLS; Mullis et al., 2012), data are collected by administering achievement tests to students of



ABOUT THE AUTHOR

Kaycheng Soh taught Educational Psychology, Research Methods and Educational Statistics at the National Institute of Education, Nanyang Technological University, Singapore. His recent research focuses on issues in international rankings and culminates in the publication of *Social and Educational Ranking: Problems and Prospects* (New York: Untested Ideas Research Centre). He served as evaluation and training consultant to many government ministries in Singapore and is currently the Research Consultant at the Singapore Centre for Chinese Language.

PUBLIC INTEREST STATEMENT

OECD's *Programme for International Student Achievement* (PISA) has caught the attention of the world's education community as well as the public. PISA compares the performance in Reading, Mathematics and Science of some 60 countries around the world. The tests are administered in the language of instruction; but, all participating countries have a sizable proportion of students who do not speak the test language at home. Understandably, this causes an underestimation of the countries' performance. This paper compares the performance in Reading of students who do and who do not speak the test language. A modified ranking system is demonstrated which takes into consideration the test language effect.

many participating countries with many different languages. As a normal practice, the tests are presented to the students in their respective language of instruction—the language officially sanctioned as the country’s medium of teaching, e.g. English in England, Finnish in Finland and German in Germany. However, as the language used for teaching and assessment may not be the home or first language of all students of a participating country, it stands to reason that the country’s performance in such international studies could have been affected, one way or another, if a large proportion of the sampled students are non-speakers of *that* language; this renders the country mean suspect as a truthful indication of the students’ performance. For instance, in PISA 2009, among the OECD countries, Iceland has the greatest proportion of 96.1% of her students being “native-born students who speak the test language at home” while, in glaring contrast, Luxembourg has only the lowest 2.6%. Likewise, among the Partner countries, Columbia has the greatest proportion of 99.3% whereas Dubai (UEA) has only the lowest 16.1%. Although care has been exercised by involving “country experts” to ensure a high degree of comparability and validity of the resultant test scores across country when used in very different cultural milieus, gross variations in the students’ linguistic background has not been taken into account in the reporting of country performance. And, this might have distorted the picture to varying extent and hence deserves research attention.

1.1. Criticisms on international testing

International testing is fraught with conceptual and technical issues although users (education officials, education-oriented politicians, etc.) of the outcomes normally take the score at face value. However, international achievement testing has been of great concerns to researchers (e.g. Mortimore, 2009; Prais, 2003; Sjøberg, 2007). In fact, PISA tests have particularly been called to question (e.g. Adams, 2003; Hassard, 2009; Mathews, 2009) with regards to their nature, format and validity.

Specifically, Prais (2003) pointed out that the kind of questions asked in the PISA were deliberately made unrelated to the school curriculum in contrast to other international studies (e.g. TIMSS), and so were unlikely to be of specific direct help to schools or to educational policy-makers (p. 152). However, in defence of PISA, Adams (2003) counter-argued that many of Prais’s criticisms were due to an incomplete understanding and knowledge of the methodology of international studies, PISA in particular.

More specifically, Hassard (2009), focusing on the PISA Science test, commented that it purported to assess scientific knowledge and use of that knowledge, understanding of the characteristic features of science, awareness of how science and technology shape and the world’s willingness to engage in science-related issues (para. 3). But, Hassard further commented that according to many authors, PISA had developed an assessment system that aligns “very well” with Vision II (which emphasizes science in life situations in which science plays a key role); but, in his own view, the PISA test is simply another large-scale test that really does not assess how students use science in lived experiences (p. 8).

In contrast, Mathews (2009) focused on the PISA Mathematics test and cited international testing expert Loveless who agreed that the example problems from the PISA test were ill-chosen and said that the test would throw kids off as the Math was rather trivial. Mathews further pointed out that there were other problems with PISA, such as an ideological effect and a tendency to assume cause-and-effect relationships (para. 4).

Likewise, Mortimore (2009) considered PISA as having disregard for national curriculum, as the tests had emphasis on asking questions which could be answered using common sense rather than knowledge of a particular curriculum (p. 5). Besides, there were 15 European scholars who were sceptical of the PISA tests; they tried to engage the PISA team in a public debate but without success. Among these European researchers, Sjøberg (2007) was particularly critical of the PISA tests and challenged the wisdom of PISA’s claim to measure students’ real life experiences. This is clearly reflected in the quote below:

The main point of view is that the PISA ambitions of testing “real-life skills and competencies in authentic contexts” are by definition alone impossible to achieve. A test is never better than the items that constitute the test. Hence, a critique of PISA should not mainly address the official rationale, ambitions and definitions, but should scrutinize the test items and the realities around the data collection. The secrecy over PISA items makes detailed critique difficult (Cited in Hassard, 2009, p. 8)

1.2. Test language effect

The PISA’s concept of *literacy* has been applied to all three subjects of Reading, Mathematics and Science. Thus, what has been said about PISA Science and Mathematics tests are equally applicable to the PISA Reading test. Moreover, it has been found that the correlations among the performance for the three tests are extremely (and unusually!) high with coefficients reaching beyond $r = .95$ (Soh, 2013).

It is a truism that the language of an achievement test pre-conditions the level of performance of students who are assessed by the test via the test language. In the case of a multiple-choice test, the students need to understand the item stems as well as the options. For open-ended questions, the students need to know what is being asked and are able to write their answers correctly. Without sufficient command of the test language, the students may not respond correctly even if they actually know the correct answers. Here lies a hidden and oft-neglected problem of test language effect in PISA and its likes.

PISA, TIMSS and PIRLS involve students who speak the test language at home as well as those who do not. And, as will be shown later, the proportions of these two types of students vary from country to country. For the student sample of any participating country, the test language may be more familiar to some students who speak it at home but not the others. When speakers and non-speakers of test language are pooled as a national sample, the country mean could well be underestimated since the non-speakers of test language can logically be expected to have poorer command of the test language.

While the cited criticisms deal with many and varied aspects of the PISA tests, test language effect in terms of the students’ linguistic background, which is the focus of the present study, is obviously not one of them.

With the above as the background, the present study is an attempt to show that the Reading means reported in PISA 2009 may not truly reflect the performance of each participating country; more specifically, the Reading means have generally been underestimated due to the presence of sizable proportions of non-speakers of test language who are by default second-language speakers where the test language is concerned, since it is not the language they speak at home.

2. Method

2.1. Data

PISA 2009 has 65 participating countries, with 34 OECD Member countries/economies and 31 Partner countries/economies (hereafter *countries* for brevity). Countries which do not have full information needed for this analysis were excluded. This left 28 OECD and 28 Partner countries, totalling 56. The data used in the present study were gleaned from Table I.A for the original Reading country means (OECD, 2010a, p. 13) and Table II.4.4 for the proportions and Reading means of (1) native-born students speaking test language at home and (2) of those speaking another language at home (OECD, 2010b, p. 177).

PISA reported the proportion of native-born students who speak the test language at home and the native-born students who do not do so. Native students are “those students born in the country of assessment, or those with at least one parent born in that country; students who were born

abroad with at least on parent born in the country of assessment” (OECD, 2010b, p. 128). In addition, there are students with immigration background who do not fit into this definition.

The PISA Reading test is made up of three components dealing with three aspects of mental strategies, approaches or purposes that readers use to negotiate their way into, around and between texts (OECD, 2010a, pp. 42–43). The first subtest, *Access and Retrieve*, assesses skills related to finding, selecting and collecting information. The second subtest, *Integrate and Interpret*, evaluates processing of what is read to make internal sense of a text and requires the understanding of the relations between different parts of a text. The third subtest, *Reflect and Evaluate*, involves drawing on knowledge, ideas or values external to the text and reflecting on a text and relate own experience or knowledge to the text. However, the present study used on the total Reading score which is a combination of the three subtest scores so as to avoid complicating the main issue of test language effect. It is argued that whatever effect test language has on one or the other subtest is likely to happen to the other as well as Reading as whole.

2.2. Analysis

To ascertain the effect of test language, comparisons were made between the sample as a whole, the subsample of speakers of test language, and the subsample of non-speakers of test language. Effect size in terms of standardized mean difference was calculated for evaluating the magnitude of difference between groups with reference to Cohens’ (1988) criteria.

Likewise, comparisons were made for the three pairs between the original and the adjusted means to enable an evaluation of the adjustment made to the original means. The nations were then grouped in terms of gains in ranking as compared with the original rankings.

Moreover, multiple regressions were run to evaluate the contributions of the original national means for Reading and the proportion of test language speakers to the adjusted Reading means.

It is necessary to point out at the juncture that the nations participated in PISA 2009 are the units of analysis and not the individual students. This is important to avoid committing ecological fallacy as the findings at the national level may not be applicable at the student level.

3. Results

3.1. Comparisons of whole sample and subsamples

As shown in Tables 1 and 2, for the 56 participating countries as a whole, Reading has a mean of 462.5 and there are 80.1% test language speakers. This suggests that there could well be an underestimation of the Reading performance. That this is indeed the case is shown by the Reading mean of 468.5 for speakers of test language speakers and 429.3 for non-speakers of test language, with a difference of six scale points. Following the approach used in the PISA report (OECD, 2010b, p. 148), the corresponding effect size was calculated as the mean difference standardized by the average of the two variances. The standardized mean difference turns out to be $d = -.11$ which fall below $d = |.2|$, indicating a trivial effect by Cohen’s (1988) criterion. On the other hand, when the sample was compared with the non-speakers of test language, the difference of 33.2 has a medium effect size of

Table 1. Reading means and test language effect

	PISA samples	Speakers of test language	Speakers of test language	Non-speakers of test language
	Reading	%	Reading	Reading
All (N=56)	462.5 (52.1)	80.1 (21.9)	468.5 (57.5)	429.3 (56.5)
OECD (N=28)	492.6 (21.4)	83.1 (17.6)	502.5 (21.3)	459.7 (37.0)
Partner (N=28)	432.3 (56.5)	77.1 (25.5)	434.5 (62.3)	389.9 (56.6)

Table 2. Mean comparisons on reading and effect sizes

	PISA sample vs. speakers of test language		PISA sample vs. non-speakers of test language		Speakers vs. non-speakers of test language	
	Difference	Effect size	Difference	Effect size	Difference	Effect size
All (N=56)	-6.0	-.11	33.2	.61	39.2	.69
OECD (N=28)	-9.9	-.46	32.9	1.09	42.8	1.42
Partner (N=28)	-2.2	-.04	42.4	.75	44.6	.75

$d = .61$. Moreover, the difference between the speakers and non-speakers of test language is 39.2 with a medium effect size of $d = .69$.

The above findings are based on averages (means). And, it is a well-known fact that averages hide important variability. The large standard deviations accompanying the means signal the large differences which warrant attention. As shown by comparisons of data in the original Table I.A (OECD, 2010a) and Table II.4.4 (OECD, 2010b), for the sample as a whole, the country differences between the reported country means for the whole sample and test language speakers vary from -45 (Luxembourg) to 73 (Dubai) with a range of as much as 118. Between the reported country means for the whole sample and non-speakers of test language, the country differences vary from -45 (Kyrgyzstan) to 98 (Peru), with a range of 143. This indicates that test language effect benefitted some countries but disadvantaged others.

3.2. Adjusting for test language effect

If test language effect paradoxically underestimates and overestimates Reading performance at the same time, the country means reported do not truly represent performance levels of the countries, and hence need be adjusted by taking into account the proportion of test language speakers vis-à-vis non-speakers.

As the test language speakers are native-born students, the native-born non-speakers and non-native non-speakers can be pooled as a group of non-speakers of test language assuming that the non-native students are most likely also non-speakers of test languages. It is further assumed that the non-native non-speakers would perform at best on par with the native-born non-speakers. Thus, each country has a group of test language speakers and the rest are non-speakers.

With the students thus classified, an adjustment can then be made to the country means for Reading. This was done with the following formula:

$$\text{Adjusted Mean} = (P \times \text{MeanP} + Q \times \text{MeanQ}) / (P + Q)$$

where P is the number of native students who speak test language, Q is the sample size— P , MeanP is the mean of native-born speakers of test language, and MeanQ is the mean of all non-speakers of test language.

Table 3 shows the original and the adjusted Reading means for the sample as a whole and the OECD and Partner countries. As shown therein, the adjusted means are slightly higher than the original means for the sample as a whole and the OECD countries but slightly lower for the Partner countries. Although the effect sizes look trivial when seen separately, the reversal in sign is not to be neglected. Specifically, the total effect of the OECD and Partner countries is a difference of 6.4 between them. The high but non-perfect correlation of $r = .97$ between the original and adjusted country means suggest that they do not rank the countries exactly the same way; there are changes in the ranking which are studied further in details.

Table 3. Comparisons of original and adjusted means for reading

	PISA samples	Adjusted mean	Difference	Effect size
All (N=56)	462.5 (52.1)	462.8 (53.5)	-.03	-.01
OECD (N=28)	492.6 (21.4)	496.0 (20.1)	-3.4	-.16
Partner (N=28)	432.3 (56.5)	429.3 (55.7)	3.0	.05

Table 4. Gains in ranking

	All (N=56)	OECD (N=28)	Partner (N=28)
More than 10	1 (2)	1 (4)	-
6-10	2 (4)	2 (7)	-
1-5	27 (48)	13 (46)	14 (50)
No change	12 (21)	3 (11)	9 (32)
-1 to -5	8 (14)	6 (21)	2 (7)
-6 to -10	5 (9)	3 (11)	2 (7)
More than -10	1 (2)	-	1 (4)

Note: Figures in parentheses are column percentages.

3.3. Impact on ranking

The impact on ranking of adjusting the country means can be evaluated by comparing the countries' rankings based on the original Reading means and those based on the adjusted means. As shown in Table 4, for the sample as a whole, 54% of the 56 countries gained in position while 25% lost, with 6% gaining more than six positions, 11% lost more than six positions, whereas 21% neither gained nor lost.

Comparing the effect on the two sets of countries, 11% of the OECD countries gained by six or more positions with 11% losing by six or more positions. At the same time, none of the Partner countries gained by six or more position but half gained by one to five positions, and 11% lost by six or more positions.

3.4. Predicting adjusted Reading means

The above results show the need and impact of adjusting Reading country means for test language effect. A pertinent question is, *How well this can be predicted and by which predictors?* The four candidates of predictors are (1) the original sample reading means, (2) the proportion of test language speakers among the native-born students, (3) the speakers' Reading performance and (4) the non-speakers' Reading performance. To guide the choice of one or more of these as predictors, a correlation analysis was run first.

As shown in Table 5, only the original sample reading means (Predictor 1) and proportion of test language speakers (Predictor 3) correlate significantly with the adjusted Reading country means (criterion).

When these two Predictors were entered to predict Adjusted Reading country means, the results are shown in Table 6. The multiple regression coefficient is $R = .994$ with a corresponding adjusted coefficient of multiple determination $R^2 = .988$, indicating that 98.8% of the criterion variance were predicted by the two predictors. Of these, proportion of test language speakers is a more powerful predictor as the ratio of 2.7 between the two β -coefficients indicates the relative contributions to the criterion.

Table 5. Correlation coefficients

	Criterion	Predictor 1	Predictor 2	Predictor 3	Predictor 4
Criterion: Adjusted Reading	1.00	.97	.06	.99	-.10
Predictor 1: Sample Reading		1.00	-.04	.97	-.12
Predictor 2: TL-speakers (%)			1.00	-.01	-.69
Predictor 3: TL-speakers' Reading				1.00	-.05
Predictor 4: Non-TL-speakers' Reading					1.00

Note: Coefficients in bold are statistically significant ($p < .05$, two-tailed).

Table 6. Multiple regression

	<i>b</i>	S.E.	β	<i>t</i>	<i>p</i>
Sample Reading	.274	.078	.267	3.531	.001
TL-speakers (%)	.680	.070	.730	9.667	.001
Constant	17.466	9.091	-	1.921	.060

Notes: $R = .994$, Adjusted $R^2 = .988$.

4. Discussion

These observed differences, reported earlier together, show that the Reading mean for the national samples as a whole underestimates the performance of the test language speakers and, at the same time, overestimates that of the non-speakers. The difference is too large to be dismissed as inconsequential. When similar comparisons were made for the OECD and the Partner countries separately, the same trends were found. However, going by the mean differences and their corresponding effect sizes, the test language effect tends to be more severe for the OECD countries than for the Partner countries. The fact that there are vast differences between test language speakers and non-speakers in Reading means indicate that test language effect is real and is not to be dismissed; this is true for the PISA sample as a whole as well as for the OECD and the Partner countries considered separately.

The finding that more cases of gains in ranking are found for the OECD countries than the Partner countries indicates that adjusting for test language effect is more beneficial to the former than the latter countries. Why should this differential benefit occur is a topic for further study. Details of the changes can be found in Table A1 in Appendix.

5. Conclusion

This study is an attempt to bring out the problem of test language effect in international achievement comparison, using PISA 2009 Reading scores to illustrate. The test language effect is attributable to heterogeneity of language proficiency of the country samples. The proportions non-speakers of test language vary widely from country to country and hence the magnitudes of effect also vary. It is readily appreciated that where the proportions of such students are large, the effect can be expected to be severe. As shown in the Appendix, discrepancies of 10 or more scale points between the original and adjusted Reading means are found for Australia (-12), Belgium (13), Germany (13), Italy (10), Luxembourg (28), New Zealand (-18) among the OECD countries. As for the Partner countries, large discrepancies are found for Dubai (-49), Hong Kong-China (-24), Macao-china (-27) and Qatar (-43).

Thus, by *not* taking into account this influence of test language, PISA 2009 has underestimated the reading performance of some countries and, at the same time, overestimated some other countries. It is herewith believed that the adjusted Reading mean which takes into account test language effect is a more valid representation of the countries' performance levels. Whatever the magnitude and direction of the test language effect, by rendering it to a background information rather than

explicitly building it into the country scores, the PISA 2009 results do not give the intended true picture of the relative strengths of the participating countries because the reported Reading performance is compounded by students' language backgrounds.

When the participating countries are compared and ranked with a tacit assumption of homogeneity in language background (i.e. all students speak the test language at home) and, the test language effect is masked. When this is the case, as the cliché goes, apples are mixed with oranges; and, the picture created by the original Reading country means may not truly reflect the actual situation and therefore need be read with due caution.

It is therefore suggested here that the country means for Reading be adjusted by combining the scores of the test language speakers and no-speakers duly weighted by their respective proportions in the national samples. This avoids the reported country means being overestimated by one group of students *and at the same time* underestimated by the other.

Although the test language effect is found here with the PISA 2009 data analysed for the present study, it stands to reason that the same compounding effect can be detected in other international comparative studies, since they use very much the same approach to collect and analyse data and report outcome.

6. Caveats

Although the study started with a not unreasonable tacit assumption that the proportion of non-speaker of test language will have an adverse effect on Reading country means, the relation turns out to be not so straightforward; for instance, Australia has 23.7% such students and the adjusted Reading mean is 12 points lower than the original Reading mean, but Spain has 23.4% such students and the adjusted Reading mean is only lower by 7 points compared with the original mean; and, at the same time, Israel has 22.4% such students and the two means are practically identical (with negligible difference of unity). These suggest that *linear* relation between the proportion of non-speakers of test language and performance cannot be assumed; something else need be invoked to explain the discrepancies, one way or the other. One possible candidate is the effectiveness of second-language teaching in that the test language effect may be nullified or minimized when the test language ("first language" in the administrative sense or language of instruction) is taught effectively even to the non-speakers of it. Countries large proportion of non-speakers of text language such as Singapore (62.2%), Thailand (48.6%) and Indonesia (65.5%) are interesting cases as their respective original and adjusted Reading are the same for practical purpose. Obviously, this is an aspect of test language effect deserving further research effort.

Another caveat is that the present study used country as the unit of analysis and processed the country means as the "country score". This is consistent with the approach used by the PISA to compare participating countries each as a unit, but then whatever is found and said here about the countries may not be applicable at the student level. It is interesting for both methodological and practical interest if multi-level analysis is conducted to separate student effects and country effects for a deeper understanding of test language effect.

Funding

The authors received no direct funding for this research.

Author details

Kaycheng Soh¹

E-mail: sohkaycheng@hotmail.com

¹ National Institute of Education, Nanyang Technological University, Singapore, Singapore.

Citation information

Cite this article as: Test language effect in international achievement comparisons: An example from PISA 2009, K. Soh, *Cogent Education* (2014), 1: 955247.

References

- Adams, R. J. (2003, September). Response to "Cautions on OECD's recent educational survey (PISA)". *Oxford Review of Education*, 29(3). Retrieved from http://www.pisa.oecd.org/Docs/Download/adams_response_prais.pdf
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Hassard, J. (2009, November 6). *PISA: Can this test measure the outcomes of progressive science education?* Retrieved from <http://www.artofteachingscience.org/2009/11/06/pisa-can-this-test-measure-the-outcomes-of-progressive-science-education/>

Mathews, J. (2009, October). *Test that makes US look bad may not be so good*. Retrieved from http://voices.washingtonpost.com/class-truggle/2009/10/politicians_and_pundits_are_us.html

Mortimore, P. (2009). *Alternative models for analysing and representing countries' performance in PISA*. Paper commissioned by Education International Research Institute, Brussels. Retrieved from <http://download.ei-ie.org/Docs/WebDepot/Alternative%20Models%20in%20PISA.pdf>

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

OECD. (2010a). *PISA 2009 results: What students know and can do—Student performance in reading, mathematics and science (Vol. I)*. Retrieved from <http://dx.doi.org/10.1787/9789264091450-en>

OECD. (2010b). *PISA 2009 results: Overcoming social background: Equity in learning opportunities and outcomes (Vol. II)*. Retrieved from <http://dx.doi.org/10.1787/9789264091504-en>

Prais, S. J. (2003). Cautions on OECD's recent educational survey (PISA). *Oxford Review of Education*, 29, 139–163. Retrieved from <http://www.pisa.oecd.org/Docs/Download/prais.pdf>, <http://dx.doi.org/10.1080/0305498032000080657>

Sjøberg, S. (2007, October). PISA and “Real life challenges”: Mission impossible? In Hopman (Ed). *PISA according to PISA (Revised Version October 8, 2007)*. Retrieved from <http://folk.uio.no/sveinsj/Sjoberg-PISA-book-2007.pdf>

Soh, K. (2013). Finland and Singapore in PISA 2009: Similarities and differences in achievement and school management. *Compare: A Journal of Comparative and International Education*, 44, 455–471. doi:10.1080/03057925.2013.787286.

Appendix

Table A1. Test language effect at country level

	Country	Percent of native-born students who speak the test language at home	Original Reading country mean	Adjusted Reading country mean	Difference between original and adjusted means
OECD	Australia	76.3	515	503	-12
	Austria	86.2	470	478	8
	Belgium	71.8	506	519	13
	Canada	73.7	524	521	-3
	Denmark	92.1	495	499	4
	Estonia	90.2	501	502	1
	Finland	95.6	536	537	1
	France	85.7	496	499	3
	Germany	83.4	497	510	13
	Greece	90.0	483	482	-1
	Iceland	96.4	500	503	3
	Ireland	89.8	496	503	7
	Israel	77.4	474	475	1
	Italy	84.0	486	496	10
	Luxembourg	2.6	472	500	28
	Mexico	96.1	425	429	4
	Netherlands	87.6	508	516	8
	New Zealand	72.5	521	503	-18
	Norway	91.2	503	505	2
	Portugal	94.2	489	493	4
	Slovak Rep.	94.3	477	479	2
	Slovenia	91.6	483	487	4
	Spain	76.6	481	488	7
	Sweden	88.5	497	503	6
	Switzerland	75.6	501	510	9
	Turkey	95.8	464	466	2
	United Kingdom	88.3	494	493	-1
	United States	79.3	500	496	-4

(Continued)

Table A1. (Continued)

	Country	Percent of native-born students who speak the test language at home	Original Reading country mean	Adjusted Reading country mean	Difference between original and adjusted means
Partner	Albania	98.4	385	389	4
	Azerbaijan	90.0	362	363	1
	Brazil	98.6	412	415	3
	Bulgaria	89.0	429	435	6
	Chinese Taipei	77.9	495	502	7
	Colombia	99.3	413	416	3
	Croatia	88.0	476	478	2
	Dubai (UAE)	16.1	459	410	-49
	Hong Kong	57.8	533	509	-24
	Indonesia	35.5	402	403	1
	Jordan	84.0	405	405	0
	Kazakhstan	79.4	390	392	2
	Kyrgyzstan	79.6	314	318	4
	Latvia	87.1	484	486	2
	Lithuania	94.7	468	471	3
	Macao-China	24.0	487	460	-27
	Montenegro	92.4	408	407	-1
	Panama	92.4	371	380	9
	Peru	94.7	370	375	5
	Qatar	34.2	372	329	-43
	Romania	96.6	424	426	2
	Russian Fed.	80.9	459	459	0
	Serbia	89.0	442	439	-3
	Shanghai-China	98.1	556	557	1
	Singapore	37.8	526	525	-1
	Thailand	51.4	421	422	1
Trinidad and Tob	95.4	416	422	6	
Uruguay	97.3	426	429	3	



© 2014 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 3.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

