

RESEARCH

Open Access

# A quantitative synthesis of the immigrant achievement gap across OECD countries

Anabelle Andon<sup>1\*</sup>, Christopher G Thompson<sup>2\*</sup> and Betsy J Becker<sup>2</sup>

\* Correspondence:

a.andon@miami.edu;  
cgthompson@fsu.edu

<sup>1</sup>Department of Educational and Psychological Studies, University of Miami, 5202 University Drive, Coral Gables, FL 33124, USA

<sup>2</sup>Educational Psychology and Learning Systems, Florida State University, 307D Stone Building, Tallahassee, FL 32306, USA

## Abstract

**Background:** While existing evidence strongly suggests that immigrant students underperform relative to their native counterparts on measures of mathematics, science, and reading, country-level analyses assessing the homogeneity of the immigrant achievement gap across different factors have not been systematically conducted. Beyond finding a statistically significant average achievement gap, existing findings show considerable variation. The goal of this quantitative synthesis was to analyze effect sizes which compared immigrants to natives on international mathematics, reading, and science examinations.

**Methods:** We used data from the Trends in International Mathematics and Science Study (TIMSS), the Programme for International Student Assessment (PISA), and the Progress in International Reading Literacy Study (PIRLS). We investigated whether the achievement gap is larger in some content areas than others (among mathematics, science, and reading), across the different types of tests (PISA, TIMSS, PIRLS), across academic grades and age, and whether it has changed across time. Standardized mean differences between immigrant and native students were obtained using data from 2000 to 2009 for current Organisation for Economic Co-operation and Development (OECD) countries.

**Results:** Statistically significant weighted mean effect sizes favored native test takers in mathematics ( $\bar{d}_{\text{math}}^* = 0.38$ ), reading ( $\bar{d}_{\text{reading}}^* = 0.38$ ), and science ( $\bar{d}_{\text{science}}^* = 0.43$ ). Effects of moderators differed across content areas.

**Conclusions:** Our analyses have the potential to contribute to the literature about how variation in the immigrant achievement gap relates to different national-level factors.

**Keywords:** PISA; TIMSS; PIRLS; Immigrants; Quantitative synthesis; OECD countries

## Introduction

Immigration has gained increasing attention worldwide in recent years. It has steadily increased in the past five decades, primarily in developed countries (OECD 2010a). This is especially true for traditional countries of immigration, or those largely defined by a history of settlement through immigration (Buchmann & Parrado 2006) – the United States, New Zealand, Australia, Canada, and more recently countries such as Germany. In these countries, the stock of the population that is foreign-born has steadily increased since the beginning of the past decade (OECD 2010b). Immigration is a multi-faceted and complex activity. It addresses important demands of the job market, such as filling gaps created by rapidly-aging populations and decreasing fertility rates. Furthermore, it is related to issues of human rights, as immigrants tend to

migrate due to reasons of political, racial, economic, or social strife. The flow of people into a given country raises many issues, including the extent to which immigrants become successful members of society. For the youngest of immigrants, success in school is one of the most important indicators of success in society.

The present analysis uses quantitative synthesis methods to examine the extent to which the gap in achievement between immigrants and natives varies at a national level. To our knowledge, only one study has compared the magnitude of the immigrant achievement gap across content areas. Schnepf (2007) separately analyzed the three data sets we combined.

## Background

One of the most influential factors in the future success of immigrants, particularly children, is education. Internationally, evidence demonstrates that immigrant students are at an educational disadvantage, typically scoring lower on assessments of science, mathematics, and reading, leading to poor educational outcomes such as a low likelihood of participating in pre-primary education and low graduation rates (e.g., Ammermuller 2007; Heus et al. 2009; Ma 2003; OECD 2010a; Portes & MacLeod 1996; Portes & MacLeod 1999; Rangvid 2007; Rangvid 2010; Zinovyeva et al. 2008).

Immigrants' success or failure largely depends on the opportunities they encounter. International educational achievement has been an important factor for economic growth (Hanushek & Kimko 2000), yet strong evidence suggests educational opportunities are not provided equally to immigrants as they are to natives. Without adequate educational opportunities and, subsequently, adequate pay, immigrants may become a permanent part of the underclass and "foster undesirable subeconomies" to the detriment of society as a whole (Martin 1999, p. 1). Furthermore, the successful integration of immigrants is essential for the maintenance of a stable society, which cannot properly function when large minority groups such as immigrants live in a permanent marginal situation (Christensen 2004).

Some evidence indicates that immigrant students are more likely than natives of a country to attend low-quality schools (OECD 2010a). This raises important questions about the quality of education immigrants across the world receive. Due to the increased importance of immigration worldwide, a large number of studies have investigated issues such as employment and earnings outcomes, immigrant adjustment and adaptation, discrimination, and history. Relative to the expansive coverage of the aforementioned subjects, the educational achievement of young immigrants has received less attention in the literature.

In this quantitative synthesis we compute standardized mean differences comparing immigrant students to native students on mathematics, reading, and science in the three major cross-national assessments – TIMSS, PISA, and PIRLS. We use moderator analyses to assess the homogeneity of the gap across OECD countries and how its size relates on various macro-level dimensions. We examine one general research question, and four specific research questions through these analyses:

On average, is there an immigrant achievement gap?

1. Does the magnitude of the gap differ across content areas – mathematics, science, and reading?

2. Does the immigrant achievement gap vary across the three tests – PIRLS, PISA, and TIMSS?
3. Does the magnitude of the gap differ across grade and age?
4. Has the size of the immigrant achievement gap changed over time?

#### **Existing research on the immigrant achievement gap**

Some research on immigrant education has focused on the immigrant achievement gap and on investigating whether or not this gap exists across various countries. Most analyses compare immigrants' to native students' achievement, controlling for a variety of sociodemographic variables such as language spoken in the home, gender, and various proxies for poverty such as books owned in the home and parents' occupation. To a large extent, an immigrant achievement gap has been found across the board. The literature is inconsistent in its use of covariates. For example, while some studies control for race and ethnicity variables, others do not. It thus becomes challenging to generate theories about the immigrant achievement gap and makes comparisons of the size of the gap difficult. An analysis that investigates this deficit unconditionally avoids treating diverse conditional effects as if they were comparable. In some instances the gap is strongly associated with such variables – so strongly that the gap may become insignificant when these characteristics are entered in statistical models (e.g., Portes & MacLeod 1996; Warren 1996). In other studies, these variables do not seem to share variance with the gap (Driessen & Dekkers 1997), leading authors to conclude that institutional differences such as segregation across schools need to be statistically controlled in order to better understand how immigrant status affects achievement (Buchmann & Parrado 2006; Christensen 2004; Dronkers & Levels 2007; Marks 2005; Rangvid 2007; Schnepf 2007; Wößmann 2003). It thus becomes challenging to generate theories about the immigrant achievement gap and strengthens the need for an analysis that investigates this deficit unconditionally.

By and large, research has not indicated whether the immigrant achievement gap is a homogenous phenomenon across countries. Specifically, no systematic effort has yet been made to understand the phenomenon cross-nationally, considering possible sources of variation such as content area (i.e., academic subject) and type of content assessed. A systematic analysis is necessary to soundly understand the gap and requires first looking at it unconditionally, as methodologies for controlling demographic variables in published studies vary greatly and make results difficult to compile in a comprehensive manner. Our initial investigations revealed that most publications do not report the size of the unconditional achievement gap (Thompson et al. 2011), making it difficult to compare findings across studies, and to assess the size of the immigrant achievement gap. Furthermore, an overwhelming number of published articles and reports exploring this phenomenon have used data from the 2000 and 2003 PISA assessments. Other important cross-national assessments, such as TIMSS and PIRLS, seem largely absent. Therefore, while the general consensus is that an immigrant achievement gap exists, the extent to which it varies across populations based on age or the type of content assessed has yet to be examined.

The assessments most often used in the immigrant-education literature differ across several dimensions. The PISA is an assessment of mathematics, science, and reading skills of 15 year-old students while the TIMSS measures both 4<sup>th</sup> and 8<sup>th</sup> grade students

on concepts of mathematics and science. PIRLS assesses 4<sup>th</sup> grade students on concepts of reading and literacy. According to the OECD, PISA is not linked to the school curriculum (see OECD PISA Website), but rather evaluates “to what extent students at the end of compulsory education, can apply their knowledge to real-life situations and [are] equipped for full participation in society”. The central question here is whether or not students can employ what they have learned in school to situations they are likely to encounter in their daily life – what is the yield of education at or near the end of compulsory schooling? In contrast, PIRLS and TIMSS are tied to curriculum, and evaluate achievement up to a certain point in schooling (see PIRLS & TIMSS Websites). Their central aim is to evaluate student knowledge of course content that is actually taught (Hutchison & Schagen 2007). This diversity in assessment purposes and populations raises the possibility that the immigrant achievement gap reported in existing studies may vary by the age of the students, as well as the purpose of the test or the type of content assessed. As an example, because most studies of 15 year-olds employ the PISA, the average immigrant achievement gap as currently understood in PISA may not extrapolate to younger populations.

Systematic and cross-national approaches to immigrant issues have the potential to highlight different immigrant experiences as well as to reveal international trends. Portes (1997) suggests such research is useful for three specific reasons:

...first, to examine the extent to which theoretical propositions “travel,” that is, are applicable in national contexts different from that which produced them; second, to generate typologies of interaction effects specifying the variable influence of causal factors across different national contexts; third, to themselves produce concepts and propositions of broader scope. (p. 820)

Our study targets Portes’s third point. We believe this study will contribute to the growing literature on the immigrant achievement gap.

We do not include other macro-level factors as moderators, as the information available in most administrations of the three tests is limited, namely, information on the origin country of immigrant pupils is not available. Recent literature indicates that in order to fully understand immigrants’ outcomes in a destination country, we must account for both origin and destination effects, the so-called ‘double perspective’ approach (Dronkers et al. 2014; Levels & Dronkers 2008; Levels et al. 2008; see Limitations and Conclusion).

## Methods

Our quantitative synthesis did not retrieve data from publications of secondary analyses, which is the traditional mode of data retrieval for quantitative syntheses, or meta-analyses. Rather we computed mean differences directly from raw data. The primary reason for this was that, in the main, existing studies of the immigrant achievement gap did not provide the information necessary for us to compute effect sizes, which would have limited our analysis significantly. In addition, because the international databases are available free of charge from the IEA (for PIRLS & TIMSS) and OECD websites (for PISA), secondary analyses are not necessary. Using the primary datasets reduces potential sources of errors introduced when compiling data from published studies that may have used different methods of data extraction and aggregation.

Still, although data retrieval for this study was less conventional, other methods employed for the study were the same as those used in typical meta-analyses. Within content areas (i.e., mathematics, reading, and science), we obtained single effect sizes for each grade within each OECD country.

We included only OECD countries in this analysis for several reasons. First, at least a third of all immigrants across the world move from developing to developed countries or from one developed country to another (UNDP 2009). The OECD is an organization composed of some of the world's most advanced and developed countries, many of which experience significant immigration. Second, in the past decade, the OECD has devoted significant attention to the issue of immigration within its member countries. It has released yearly publications such as the *International Migration Outlook*, *Where Immigrant Students Succeed – A Comparative Review of Performance and Engagement in PISA 2003*, and *Equal Opportunities-The Labour Market Integration of the Children of Immigrants* (see for example OECD 2006a, b, c, 2010a, b). We focused on the immigrant achievement gap as a phenomenon particular to a specific type of immigration – country to country as opposed to within country migration<sup>a</sup>. The latter has not been well investigated in the immigrant achievement gap literature and cannot be studied given the data available from the three testing programs.

Effect sizes were computed with various software using data available in the original datasets for PIRLS, PISA, and TIMSS. An immigrant was defined as a student not born in the country of testing. Immigrant status is derived from a “yes” or “no” question included in all three assessments that asks the student whether or not they were born in the country of testing.

First, we computed means and standard deviations for native and immigrant students for each country using the *International Data Base (IDB) Data Analyzer* (IDB Analyzer (Version 2) 2009), an application developed by the IEA Data Processing and Research Center to be used in conjunction with SPSS. The IDB Analyzer uses the total student weight and five plausible values for each outcome (OECD 2006b; OECD 2003a; Martin & Kelly 1996; Martin et al. 2003) to obtain population estimates of mean performance as well as an estimate of the variance of this quantity at the country level. We then computed mean differences, effect sizes, and effect-size variances using Excel because the IDB analyzer does not compute effect sizes or effect-size variances. All other calculations and analyses were conducted in R (R Development Core Team 2011; R Core Team 2014) using the *metafor* package (Viechtbauer 2010a, b).

### Moderators

Moderators are variables that may affect or relate to the sizes of effects. The three moderators in this study (*year*, *test*, and *grade*) were selected according to gaps in current research. First, as previously discussed, current knowledge based on studies of the immigrant achievement gap may only be generalizable to 15-year olds tested on concepts attached to real-world applications. Thus, we examined the test – PISA, TIMSS, or PIRLS – and related features – year implemented and grade assessed – as moderators. We investigated whether the immigrant achievement gap has changed across time in the past decade, considering that most existing studies have only employed the 2000 and 2003 PISA data. In regression analyses moderators *test* and *grade* were treated as discrete variables<sup>b</sup> and *year* was continuous and centered at 2000 (the first year of data).

### Effect sizes

To quantify the immigrant achievement gap, we calculated mean differences at the country level between immigrant and native test takers. Considering differences among measures and scales for outcomes, a standardized mean difference was the most reasonable choice for computing effect sizes given the aggregated format of the data. The unbiased standardized-mean-difference effect size is

$$d_i = \left( 1 - \frac{3}{4(n_i^N + n_i^I) - 9} \right) \frac{\bar{Y}_i^N - \bar{Y}_i^I}{S_i}, \quad (1)$$

where  $\bar{Y}_i^N$  and  $\bar{Y}_i^I$  are the sample mean outcomes for the respective native and immigrant samples of the  $i^{\text{th}}$  test administration,  $n_i^N$  and  $n_i^I$  are the respective native and immigrant sample sizes from the  $i^{\text{th}}$  test administration, and  $S_i$  is the pooled standard deviation of the  $i^{\text{th}}$  sample (Hedges 1981)<sup>c</sup>. Therefore, a positive effect size is interpreted as an achievement gap which favors native test takers, and a negative effect size favors immigrant examinees. As shown in Hedges (1981) and Borenstein (2009, p. 226), the variance of  $d_i$  is

$$v_i = \frac{n_i^N + n_i^I}{n_i^N n_i^I} + \frac{d_i^2}{2(n_i^N + n_i^I)}. \quad (2)$$

Data were gathered for all TIMSS, PISA, and PIRLS administrations during the years 2000 to 2009 for all countries that were members of the OECD as of 2011. This resulted in an initial set of 542 unique effect sizes classified by *country*, *year*, *test*, *grade*, and *content area*. Except for when comparing across content areas, samples for which effect sizes were computed are independent. Typically, during any given test administration, students complete tests in multiple content areas at one time. This results in a dependency in responses across content areas, which further translates to dependence among effect sizes. To address this issue, we analyzed the three content areas separately.

Some test administrations had very low immigrant sample sizes, the lowest of which was only two students. Because the consistency and efficiency properties of the standardized mean difference rely on large sample statistical theory, we excluded samples which had an immigrant sample size ( $n^I$ ) less than 30. As a result, 29 effect sizes were excluded (roughly 5% of the original sample), bringing the number of effect sizes used in the quantitative synthesis to 513.

### Analyses

A typical method of choosing a quantitative synthesis model (fixed or random effects) is to determine the extent of homogeneity among effect sizes. Multiple methods have been proposed, the most common being the homogeneity test referred to as the  $Q$  test. The formula for  $Q$ , as shown in Shadish and Haddock (2009), is

$$Q = \sum_{i=1}^k v_i^{-1} \left( d_i - \frac{\sum_{i=1}^k d_i v_i^{-1}}{\sum_{i=1}^k v_i^{-1}} \right)^2. \quad (3)$$

If all studies are homogeneous and share a common effect size,  $Q$  will be approximately distributed as a chi-square distribution with  $k - 1$  degrees of freedom ( $df$ )

(Hedges 1992). The null hypothesis tested by the  $Q$  statistic is that all effect sizes are homogenous and any variability results from sampling error. Large values of  $Q$  suggest that our collection of effect sizes is heterogeneous. Three  $Q$  statistics – one for each content area – are presented in Table 1.

A secondary index for analyzing effect-size homogeneity is the  $I^2$  index, which “... describes the percentage of total variation across studies that is due to heterogeneity rather than chance” (Higgins et al. 2003, p. 558). We calculate  $I^2$  as

$$I^2 = \frac{100(Q-k+1)}{k-1} \% \tag{4}$$

Higgins et al. (2003) interpret  $I^2$  values as showing *no variation*, *low variation*, *moderate variation*, and *high variation* for cutoffs of 0%, 25%, 50%, and 75%, respectively. As with results from the  $Q$ test, resulting  $I^2$  values (see Table 1) indicate that random-effects estimation would be appropriate.

The random-effects estimate of the mean can be interpreted as an *average* effect size because it does not assume the population of effect sizes can be completely explained by a unique effect-size representation. Among many other sources, Hedges and Vevea (1998, p. 493) present a general formula for calculating the random-effects mean effect size as

$$\bar{d}^* = \frac{\sum_{i=1}^k w_i^* d_i}{\sum_{i=1}^k w_i^*} \tag{5}$$

where  $w_i^*$  is the random-effects weight and is calculated as  $(v_i + \hat{\tau}^2)^{-1}$ . The  $v_i$  term is given in (2). The addition of  $\hat{\tau}^2$ , typically referred to as the between-studies variance, represents the presence of true variability among studies beyond sampling error. In place of the term ‘between-studies variability’ commonly used in meta-analysis applications, we will refer to between-effects variability. The between-effects variance component must be estimated; we used the commonly implemented DerSimonian and Laird (1986) estimator

$$\hat{\tau}^2 = \max \left( 0, \frac{Q-k+1}{\sum_{i=1}^k v_i^{-1} - \left[ \frac{\sum_{i=1}^k v_i^{-2}}{\sum_{i=1}^k v_i^{-1}} \right]} \right) \tag{6}$$

Last, the conditional variance of the random-effects mean is

$$v^* = \frac{1}{\sum_{i=1}^k w_i^*} \tag{7}$$

**Table 1 Homogeneity indices**

Quantity	Mathematics	Reading	Science
$Q$	7572.9*	5607.1*	6462.2*
$I^2$	97.7	97.2	97.3

Note. The degrees of freedom for the  $Q$  statistic are the same for Mathematics and Science ( $df = 176$ ) but differ for Reading ( $df = 158$ ).

\* $p < .05$ .

Using (5) and (7), a 95% confidence interval about the random-effects mean can be formed as

$$\bar{d}^* \pm 1.96\sqrt{v^*}, \tag{8}$$

and a 95% prediction interval can be calculated as

$$\bar{d}^* \pm 1.96\sqrt{\hat{\tau}^2}. \tag{9}$$

In an effort to explain between-effects variability, we examined mixed-effects regression models. These models incorporate regression coefficients that associate study characteristics (i.e., moderators) to study outcomes while allowing for unexplained variance in the model (Raudenbush 2009). Our mixed-effects regression models consider effect sizes as outcomes, and study characteristics (such as *test*) as moderators of the variability among effect sizes. For each we provide a *Q-model* statistic, denoted as  $Q_M(df)$ . This statistic assesses the amount of total variation explained by the model. When effect sizes are well-explained by the moderators,  $Q_M$  will be large. We also provide a *Q-error* statistic, denoted as  $Q_E(df)$ . This statistic assesses the amount of total variation not explained by the predictions when a fixed effects model (with explained variation not incorporated) is examined; lower  $Q_E$  values are desired. Result for  $Q_M$  and  $Q_E$  can be found in Tables 2, 3 and 4.

## Results

### Overall analyses

Figure 1 provides error bar plots for all effects by content area and shows that the ranges of effects within content areas were fairly similar. The lowest effects were medium in magnitude and negative, representing cases where immigrants outperformed natives, while the highest effects were large and positive. Across all content areas, over 80% of effect sizes were positive, indicating an achievement gap which favored native test takers. Furthermore, across all content areas no large negative effects were seen. Last, over 75% of the effects were statistically significant at the  $\alpha = .05$  level.

Table 1 provides effect-size homogeneity information for each content area. All three data sets had statistically significant *Q* statistics. In addition, the average  $I^2$  across the content area was 97%, which is very large. Both homogeneity indices agree that effect sizes for all three outcomes are heterogeneous. As previously stated these indices also indicate the appropriateness of adopting a random-effects model.

**Table 2 Mathematics mixed-effects model**

Moderator	Excluding test <sup>a</sup>			Excluding grade <sup>b</sup>		
	$\beta$	SE	<i>p</i>	$\beta$	SE	<i>p</i>
[intercept]	0.635*	0.077	< .05	0.504*	0.063	< .05
Year	-0.007	0.009	.44	-0.005	0.009	.57
Test (PISA)	—	—	—	-0.152*	0.054	< .05
Grade						
Grade 4 vs. Grade 8	-0.237*	0.085	< .05	—	—	—
Grade 4 vs. Age 15	-0.274*	0.069	< .05	—	—	—

<sup>a</sup> $Q_M(3) = 16.3^*$ ,  $Q_E(173) = 6864.6^*$ .

<sup>b</sup> $Q_M(2) = 8.0^*$ ,  $Q_E(174) = 7393.4^*$ .

\* $p < .05$ .



**Table 3 Reading mixed-effects model**

Moderator	Excluding test <sup>a</sup>			Excluding grade <sup>b</sup>		
	$\beta$	SE	<i>p</i>	$\beta$	SE	<i>p</i>
[intercept]	0.510*	0.54	< .05	0.510*	0.054	< .05
Year	-0.017*	0.008	< .05	-0.017*	0.008	< .05
Test (PISA)	—	—	—	-0.073	0.053	.18
Grade						
Grade 4 vs. Age 15	-0.073	0.054	.17	—	—	—

<sup>a</sup> $Q_M(2) = 7.8^*$ ,  $Q_E(156) = 5358.2^*$ .

\* $p < .05$ .

Table 5 provides the between-effects variances, as well as random-effects means and their associated 95% confidence and prediction intervals for all content areas. Mean effect sizes for mathematics and reading data were identical (both equal to 0.38). This indicates a small-to-moderate overall effect favoring native students. The mean effect for the science data was slightly larger (0.43) and also favored native students. All means were statistically different from zero. Last, the between-effects variances ( $\hat{\tau}^2$ ) were fairly large and similar across all subjects.

#### Mixed-effects regression models

As part of the mixed-model analyses, all data sets were checked for multicollinearity among moderators. Bivariate correlations were calculated for all predictors (see Table 6). In all three content areas the largest correlation, by far, was between *grade* and *test*. This occurs because PISA is given only to 15 year olds, PIRLS to 4th graders, and TIMSS at multiple grade levels. For this reason, we did not include *grade* and *test* simultaneously as moderators in the models. Beyond this high degree of multicollinearity, other moderators had low degrees of dependence as determined by moderately-low bivariate correlations and low variance inflation factors (all were less than two).

Tables 2, 3 and 4 provide regression coefficients, standard errors, and probability values for both models (excluding either *grade* or *test*) in each of the three content areas.

#### Mathematics data

Results for mathematics data differed based on whether grade or test was used as a moderator in the model. When *test* and *year* were modeled (i.e., excluding *grade*), the only statistically significant moderator of the size of the immigrant gap was *test*

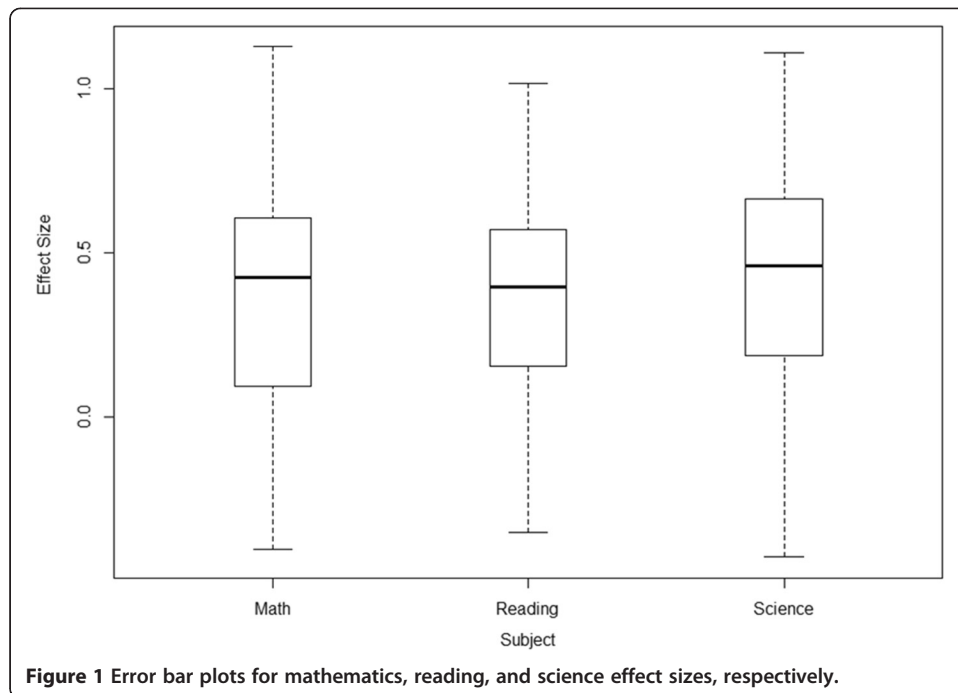
**Table 4 Science mixed-effects model**

Moderator	Excluding test <sup>a</sup>			Excluding grade <sup>b</sup>		
	$\beta$	SE	<i>p</i>	$\beta$	SE	<i>p</i>
[intercept]	0.654*	0.071	< .05	0.563*	0.058	< .05
Year	-0.004	0.008	.65	-0.003	0.008	.76
Test (PISA)	—	—	—	-0.183*	0.050	< .05
Grade						
Grade 4 vs. Grade 8	-0.166*	0.078	< .05	—	—	—
Grade 4 vs. Age 15	-0.268*	0.063	< .05	—	—	—

<sup>a</sup> $Q_M(3) = 18.9^*$ ,  $Q_E(173) = 5787.3^*$ .

<sup>b</sup> $Q_M(2) = 13.8^*$ ,  $Q_E(174) = 6099.6^*$ .

\* $p < .05$ .



( $\hat{\beta}_{\text{test}} = -0.152$ ). This implies that, holding *year* constant<sup>d</sup>, the average difference among effect sizes between the PISA and TIMSS data was 0.152. Specifically, the immigrant achievement gap is 0.152 standard deviations larger for TIMSS data than for PISA data.

When *grade* and *year* were included as moderators (i.e., excluding *test*), results were similar for the slopes representing the moderator *grade* ( $\hat{\beta}_{\text{grade [1]}} = -0.237$  and  $\hat{\beta}_{\text{grade [2]}} = -0.274$ ). The predicted size of the gap for 4<sup>th</sup> graders was 0.64 standard deviations, controlling for year. The immigrant achievement gap in math was 0.237 standard deviations smaller for 8<sup>th</sup> grade test takers than for 4<sup>th</sup> grade test takers, and 0.274 standard deviations smaller for 15 year olds than for 4<sup>th</sup> graders. Taking the difference between these slopes gives the difference between gaps for 8<sup>th</sup> graders and 15 year olds, which is a negligible 0.001 standard deviations. These values show that the size of the immigrant achievement gap is lower for all older examinees, by about one-fourth of a standard deviation.

Both models explained a significant amount of heterogeneity in the math gaps, as indicated by  $Q_M(3) = 16.3, p < .05$  and  $Q_M(2) = 8.0, p < .05$ , respectively. However, both  $Q$ -error statistics ( $Q_E$  from the fixed model) were quite large and statistically significant (see Table 2), indicating much effect-size variability has yet to be explained.

**Table 5** Random-effects means

Subject	Mean effect size ( $d^*$ )	Between-effects variance ( $\hat{\tau}^2$ )	95% Confidence interval for the mean	95% Prediction interval
Mathematics	0.38	0.11	[0.33, 0.43]	[-0.27, 1.03]
Reading	0.38	0.09	[0.33, 0.43]	[-0.21, 0.97]
Science	0.43	0.09	[0.38, 0.48]	[-0.16, 1.02]

**Table 6 Predictor correlations**

Content Area	Variable	Year	Test	Grade
Mathematics	Year	1.00	.09	-.10
	Test	.09	1.00	-.92*
	Grade	-.10	-.92*	1.00
Reading	Year	1.00	-.11	.11
	Test	-.11	1.00	-1.00*
	Grade	.11	-1.00*	1.00
Science	Year	1.00	.09	-.10
	Test	.09	1.00	-.92*
	Grade	-.10	-.92*	1.00

\* $p < .05$ .

### Reading data

In contrast to the results for mathematics, results for the reading data did not significantly differ based on whether *grade* or *test* were entered in the model. In both instances *year* was a significant moderator ( $\hat{\beta}_{\text{year}} = -0.017$ ). On average, the immigrant achievement gap has decreased by 0.017 standard deviations each year since 2000. This result is best interpreted as a weak, general trend over time rather than a year-to-year difference because none of the examinations studied is offered every year. We examine this result more closely with a cumulative quantitative synthesis in Appendix B. The reading model explained a significant amount of effect-size heterogeneity even given a large degree of uncertainty ( $Q_M(2) = 7.8, p < .05$ ). This  $Q_M$  result was the same for both *grade* and *test* models. As with the mathematics models, the  $Q_E$  error statistic ( $Q_E$ ) was quite large and statistically significant (see Table 3), which means a large degree of effect-size variability was not explained by the predictors.

### Science data

For science, the significance of the moderators in both models (i.e., with *grade* or *test*) was similar. Both *test* and *grade* explained a significant amount of effect-size heterogeneity. The slope for *test* ( $\hat{\beta}_{\text{test}} = -0.183$ ) reveals that the average effect size was 0.183 larger for TIMSS than PISA. In the case of *grade* ( $\hat{\beta}_{\text{grade [1]}} = -0.166$  and  $\hat{\beta}_{\text{grade [2]}} = -0.268$ ) results were similar to those for the mathematics data. The immigrant achievement gap was 0.166 standard deviations larger for 4<sup>th</sup> grade test takers than for 8<sup>th</sup> grade test takers, and 0.268 standard deviations larger for 4<sup>th</sup> graders than for 15 year olds. Taking the difference between these slopes gives the difference between gaps for 8<sup>th</sup> graders and 15 year olds, which is about 0.10 standard deviations. Given the intercept of 0.65, these results suggest that the immigrant achievement gap is greatest in grade 4, is about 25 percent lower for 8<sup>th</sup> graders and another 16 percent lower for the 15-year olds. Both science models explained a significant amount of effect-size heterogeneity, as respectively indicated by  $Q_M(3) = 18.9, p < .05$  and  $Q_M(2) = 13.8, p < .05$ . As with the mathematics and reading models, both  $Q_E$  error statistics ( $Q_E$ ) were quite large and statistically significant (see Table 4).

### Model fit

We also tested a series of assumptions for each linear model. First, many potential influential points had been eliminated by virtue of their small sample size of immigrant students (5% of the total set of effects were excluded, as previously mentioned). Many of the excluded effects were large. Leverage plots were also examined to determine if any influential points existed. Within the remaining effect sizes, several potential influential points were located, but their influence was minimal based on information derived from the leverage plots. Ultimately we did not exclude any additional observations as these potentially-influential points are likely not products of measurement imprecision (see our previous discussion on excluding data from small samples).

Normal quantile-quantile plots confirmed approximate normality of residuals for all content areas, and partial residual plots confirmed approximate linearity of continuous predictors related to effects, in all content areas. All preliminary assumption checks were completed using the *car* package in R (Fox & Weisberg 2011). As in all modeling scenarios, model fit can always be improved. First, though all models explained a significant amount of variability in effects (as shown by  $Q_M$ ), all model fit tests ( $Q_E$  results) were very large and statistically significant. This excessive unexplained variability may be explainable if we were to test other moderators (see Limitations). Second, variance-explained values for all six models, denoted as  $R_{meta}^2$ , were all small, ranging from almost zero to .08 (values are not shown here), further indicating the potential for other moderators to explain effect-size variability. This measure compares the variability explained by the model with no moderators to the variability explained by a model with moderators (see Aloe et al. (2010) for more information). Both of these indicators of model fit suggest further variation remains in all three sets of content-area effect sizes.

### Discussion

We found significant overall mean effect sizes for mathematics ( $\bar{d}_{math}^* = 0.38$ ), reading ( $\bar{d}_{reading}^* = 0.38$ ), and science ( $\bar{d}_{science}^* = 0.43$ ), all of which are moderate in magnitude. Prediction intervals suggested that the bulk of the effects in all areas are likely positive, favoring native students. Only 8 percent (for science) to 13 percent (for mathematics) of effects are likely to be below zero. This addresses the overarching research question and indicates that, in fact, an immigrant achievement gap exists for all assessed content areas in favor of native students. The gap for science is slightly larger than the mathematics and reading gaps, which are empirically identical. While a difference across content areas has never been previously tested with meta-analytic methods, other authors have posited such a pattern. For example, Schnepf (2007) argued that the gap would likely be larger for reading than mathematics because assessments of mathematics require fewer linguistic skills than reading assessments; this would relate directly to immigrant students' proficiency in the language of testing.

This quantitative synthesis does not completely support this notion; rather, it suggests that immigrant students are at an equal disadvantage in reading and in mathematics when compared to native students. Yet the logic presented by Schnepf (2007) may explain the significantly higher gap in science. Perhaps the language used in mathematics is more universally understood, while context in both math and reading assessments

may aid immigrant test takers who are non-native language speakers in deriving meaning in order to successfully respond to questions. Further, the content in a mathematics assessment is often numerical, and to the extent that the immigrant students' native countries use the same number system as the country of testing, this type of assessment may be less daunting than a science assessment. Unlike mathematics items, science items may tend to be word problems that include technical language in the language of testing. They also may not provide as much context as a reading passage. For example, immigrant students who do not speak the language of testing well may be able to create meaning from the reading passage. In other words, not knowing the meaning of some words may not be as detrimental when the item is longer and has context as opposed to when the item is short and lacks context or includes technical terms (which may be likelier for science items). However, such an explanation only applies to non-native language speakers. Some immigrants do speak the language of the test as a first or additional language. Perhaps this finding also hints at potential differences in quality of science curriculum and instruction between origin and destination countries. If immigrant students have been exposed to a poorer quality science instruction in their native countries, for example, then this may exhibit itself in a science immigrant achievement gap on assessments given in the destination country.

Six separate regression models, two for each content area, addressed our subsequent research questions. While statistical significance of the moderators varied, some similarities were found across the models, specifically between mathematics and science. The achievement gap was larger in TIMSS than PISA for both mathematics and science by about one to two tenths of a standard deviation. The gap was also smaller for older immigrant children by about two tenths of a standard deviation in both math and science.

Next, only one moderator was significant for the reading effects – *year* of testing. Although this quantitative synthesis is in the main cross-sectional, the significance of this moderator would indicate a possible weak trend in which the gap in reading has decreased from the beginning to the end of the last decade (see Appendix B for a slightly different perspective on the matter).

Because few studies have examined macro-level differences in the immigrant achievement gap, it is difficult to make strong theoretical interpretations of the findings. Perhaps the most significant findings are the differences across grades and tests. The fact that younger students show a larger immigrant achievement gap is not necessarily intuitive, since it is commonly believed that young children adapt to new environments more easily and learn new languages more quickly than older students. The difference we found may reflect the composition of student populations in later grades, which include those who have not dropped out of school or who have the means and the support at home to stay in school, and are thus possibly the most advantaged in a given country. This may imply that academic differences between native and immigrant students at the highest levels of privilege are still present but narrower, although our data are not disaggregated to a level where analysis of this hypothesis is possible.

The difference in the gap magnitude between TIMSS and PISA may be in part due to the type of content assessed. Specifically, TIMSS assesses the effectiveness of the curriculum whereas PISA evaluates the extent to which pupils at the end of compulsory schooling can apply what they have learned to situations they will likely encounter in

their daily lives. Content assessed in TIMSS evaluates formal mathematics knowledge, whereas items in PISA are more applied in nature as they pose real-world scenarios that require mathematics. Perhaps immigrant students fare better on items that tell a story, provide more context, and allow them to apply their experience and knowledge, such as those in PISA. Coupled with the finding that older immigrant children exhibit a narrower gap, this may indicate that immigrant adolescents who have not yet dropped out of school are nearly as ready for the workforce (as measured by PISA) as native students. Our findings seem to suggest larger disparities between younger and older students when assessed with TIMSS than PISA.

### Limitations

Several considerations suggest the use of caution when making inferences from this analysis. First, we were limited because most administrations of the three assessments did not collect country of origin information from immigrant pupils. For this reason, we could not investigate macro-level characteristics of the countries in this study. The most recent research in this area indicates that both origin and destination macro-level variables must be investigated to fully understand the immigrant achievement gap (Levels & Dronkers 2008; Levels et al. 2008; Dronkers et al. 2014). Second, the generalizability of this study is limited to OECD countries, although our initial investigations also found an overall significant mean immigrant achievement gap with a wider set of countries (Thompson et al. 2011). Third, because we defined immigrants as students not born in the country of testing, we are studying by definition only first-generation immigrants (Rumbaut 2004). Fourth, in the three testing programs, countries are permitted to exclude students who are non-native speakers of the testing language *and* who have received less than one year of instruction in that language. This study, as any other employing data from the PIRLS, PISA, and TIMSS, is representative of students who have a certain degree of proficiency in the language of testing. Fifth, some of the variation in effects found across test content may be due to the differing methodologies employed in PISA and TIMSS for calculating variance rather than an observed effect in the population. Finally, our quantitative synthesis examined the extent to which the immigrant achievement gap varied by subject. To address such a question, we compared reading, science, and mathematics scores that are not on the same scale, although standardized effect sizes in part address this issue.

We have suggested reasons for possible gap differences using several moderators. Although characteristics of an immigrant student, such as their non-native language speaker status, may contribute to the existence of a gap, they are most certainly not the only source, as previously discussed. Strong evidence has shown inequities in the quality of the education that immigrants are provided in destination countries (e.g., Conchas 2001; Crul & Holdaway 2009; Lee 2002; Minoiu & Entorf 2005; OECD 2010a; Schneeweis 2006). Although immigrant students may be at an academic disadvantage due to their individual characteristics, such as socioeconomic status and native language, the experiences they have had in both their origin and destination countries have an effect on the immigrant achievement gap. Finally, as we did not analyze student-level data, we did not investigate any student or school correlates of the immigrant achievement gap. Thus, it is difficult to conclusively discuss all possible sources of the gap. In the future, malleable factors must be investigated in order to better

understand how to close the gap. More than likely, factors found at the school level will have the most potential for reducing or eradicating this deficit.

## Conclusions

One of the aims of this quantitative synthesis was to examine the extent of the homogeneity of the immigrant achievement gap from a macro-level perspective. We found that the immigrant achievement gap is a very heterogeneous phenomenon and varies by grade and type of content assessed. It also varies by year (for reading). Thus even though gaps are present on average, they are not constant across all conditions and groups of students. In a small percent of populations, the gaps favor immigrants. Intuitively, the size of the science gap in comparison to the reading and mathematics gaps may make sense. Science assessments may include more complex and technical language than mathematics and reading assessments. Future research should investigate the content of the assessments as well as include item-level analyses in order to better understand what features of mathematics and reading assessments yield a smaller immigrant achievement gap than science assessments. The same applies to the type of content assessed in PISA and TIMSS, as evidence presented here suggests immigrants perform less poorly on PISA than TIMSS (relative to natives).

Most analyses to date have questioned whether or not a gap exists across countries, often controlling for student-level variables such as race, ethnicity, level of poverty, and native language. Our analysis demonstrates that, on average, there is a gap for the three core content areas across countries. Importantly, single-level analyses that control for student-level variables cannot answer all questions about what may explain the immigrant achievement gap. Because the gap is not a student-level phenomenon, in that no individual student him or herself can exhibit a gap, future questions about the sources of this deficit must analyze the gap as a school-level phenomenon. Further, Dronkers et al. (2014) emphasize that “contextual features of both origin and destination countries do affect the educational performance of migrant children, and must be part of any explanation of migrant children’s school success.” (p. 2). Immigrants do not arrive in destination countries as a blank slate. Factors such as their educational experiences and reasons for migration influence their degree of success in the destination country. Characteristics of the origin country such as political stability, level of economic development, and length of compulsory education have shown significant effects on the educational achievement of immigrants in the destination country (Levels & Dronkers 2008; Levels et al. 2008; Dronkers et al. 2014). To this end, future studies should continue to investigate possible moderators of the immigrant achievement gap at a national level from both origin and destination countries.

This article provides the most systematic investigation of the immigrant achievement gap to date based on three critical databases. Our analyses investigate correlates of the gap at a macro level. Our findings are consistent with the existing literature which has continuously reported an immigrant achievement gap. Our findings may allow researchers to now focus on investigating malleable factors to address this academic deficit between immigrant and native students instead of continuing to focus on whether or not a gap exists between these students. We hope that our results provide aid organizations with evidence on what variables are associated with the gap so they can tailor interventions to ameliorate the immigrant achievement gap at a national level. Future

research should begin to identify further malleable factors at the school and country levels in order to address the academic deficit between immigrant and native students.

### Endnotes

<sup>a</sup>According to the United Nations Development Programme, almost four times as many people move within countries as across countries (UNDP 2009).

<sup>b</sup>For *test*, PISA was coded as “1” and TIMSS and PIRLS were coded as “0.” A third code was not necessary because TIMSS and PIRLS data were never analyzed together because different participants are tested in the two programs. For *grade*, we created dummy variables for 4<sup>th</sup> graders (reference group), 8<sup>th</sup> graders, and 15-year olds.

<sup>c</sup>The standard deviation is  $S_i = \sqrt{\frac{(n_i^N - 1)(S_i^N)^2 + (n_i^I - 1)(S_i^I)^2}{n_i^N + n_i^I - 2}}$ , where  $S_i^N$  and  $S_i^I$  are the respective standard deviations of the native and immigrant samples for the  $i^{\text{th}}$  sample.

<sup>d</sup>Henceforth we will not repeat the phrase “holding all other moderators constant” for the sake of brevity.

### Appendix A List of OECD countries in quantitative synthesis

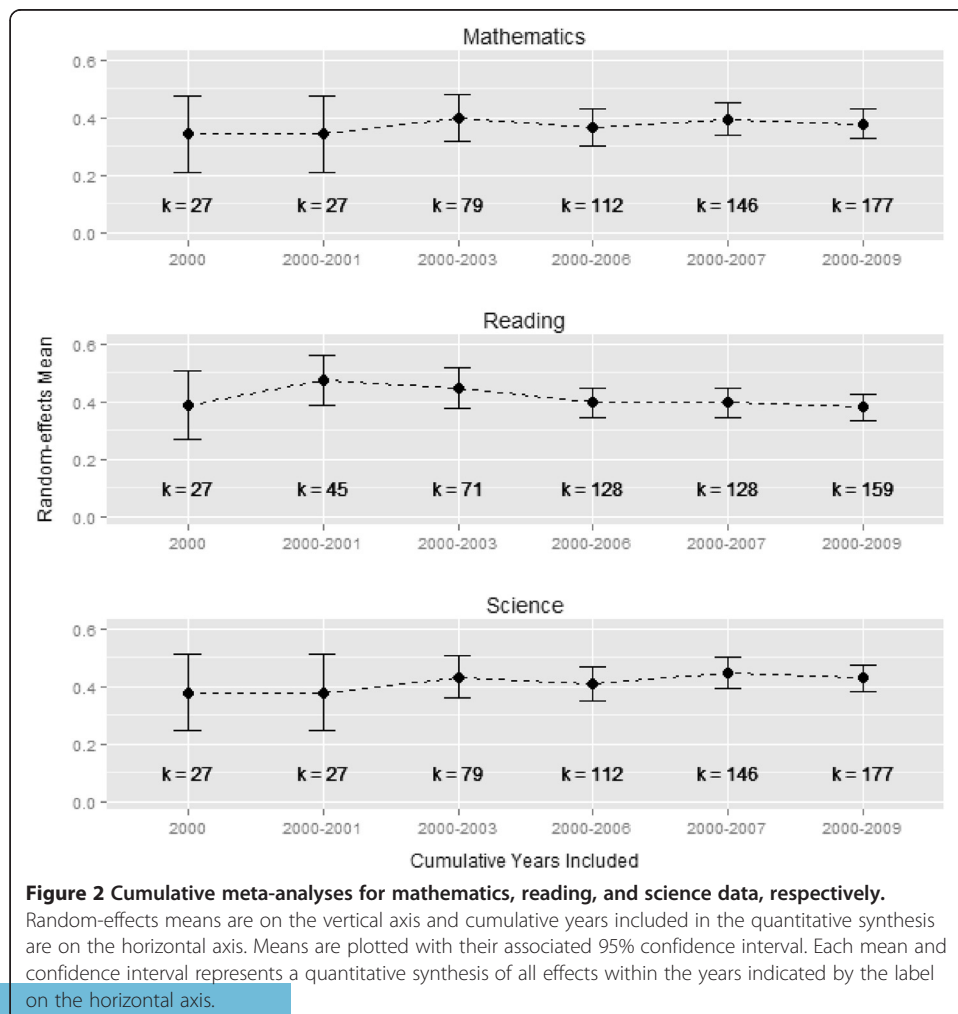
1. Australia
2. Austria
3. Belgium
4. Canada
5. Chile
6. Czech Republic
7. Denmark
8. Estonia
9. Finland
10. France
11. Germany
12. Greece
13. Hungary
14. Iceland
15. Ireland
16. Israel
17. Italy
18. Japan
19. Korea
20. Luxembourg
21. Mexico
22. Netherlands
23. New Zealand
24. Norway
25. Poland
26. Portugal
27. Slovak Republic
28. Slovenia



- 29. Spain
- 30. Sweden
- 31. Switzerland
- 32. Turkey
- 33. United Kingdom
- 34. United States

### Appendix B Cumulative meta-analyses

While investigating *year* as a predictor, we became interested in how mean effects varied over time for each content area. Therefore we completed cumulative meta-analyses for each subject. Cumulative meta-analyses include multiple, successive meta-analyses for each time point (in our case, *year*) of data. For example, our data begins at year 2000. At the first time point, *only* effects based on tests given in 2000 were meta-analyzed using the random-effects procedures described above. Next, the following time point (i.e., *year* = 2001) is considered and the same process is completed using effects from 2000 *and* 2001. This process is then repeated for all time points through 2009. The main advantage of performing a cumulative meta-analysis is the ability to



**Figure 2** Cumulative meta-analyses for mathematics, reading, and science data, respectively.

Random-effects means are on the vertical axis and cumulative years included in the quantitative synthesis are on the horizontal axis. Means are plotted with their associated 95% confidence interval. Each mean and confidence interval represents a quantitative synthesis of all effects within the years indicated by the label on the horizontal axis.

see the stabilization (or lack thereof) of mean effects over time (here, across years of testing).

Figure 2 provides the cumulative meta-analyses for all content areas. As time progresses, confidence intervals typically decrease in size, implying a more precise mean estimate. This is expected as, over time, the number of effects used to calculate the mean increases. However, in a few instances, going from one year to the next,  $k$  did not change because the given subject was not tested between those years but other(s) were. These duplicate points were nonetheless included to ensure comparability across the three plots.

Overall, results for all content areas showed fairly stable mean effects, suggesting the gap has been fairly consistent over the period from 2000 to 2009. This is confirmed by the overlap of the confidence intervals across all years, for each subject. One exception may be for the reading data, where a practically significant jump (i.e., an increase in the gap) of about one-tenth of a standard deviation was seen from 2000 to 2001. This reflects the weak, but statistically significant effect of the *year* moderator for the reading model. Practically speaking, this may mean that although the gap in reading increased between the years 2000-2001, it stabilized over time. This initial jump followed by subsequent decreases may have manifested itself in a negative effect in the reading model when, in fact, the gap was consistent across the last decade for reading as it was for science and mathematics. From a policy standpoint, this suggests that efforts to address the deficit between immigrant and native students in the core subjects have not closed the achievement gap in the past decade.

#### Abbreviations

PISA: Programme for International Student Assessment; TIMSS: Trends in International Mathematics and Science Study; PIRLS: Progress in International Reading Literacy Study; OECD: Organisation for Economic Co-operation and Development; UNDP: United Nations Development Programme; IDB: International Data Base; CIA: Central Intelligence Agency; df: Degrees of freedom.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

All authors contributed equally to the manuscript. All authors read and approved the final manuscript.

Received: 24 June 2013 Accepted: 19 June 2014

Published online: 06 August 2014

#### References

- Aloe, AM, Becker, BJ, & Pigott, TD. (2010). An alternative to  $R^2$  for assessing linear models of effect size. *Research Synthesis Methods*, 1(3-4), 272-283.
- Ammermuller, A. (2007). Poor background or low returns? Why immigrant students in Germany perform so poorly in the Programme for International Student Assessment. *Education Economics*, 15(2), 215-230.
- Borenstein, M. (2009). Effect sizes for continuous data. In HM Cooper, LV Hedges, & JC Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 221-235). New York: Russell Sage.
- Buchmann, C, & Parrado, EA. (2006). Educational achievement of immigrant-origin and native students: A comparative analysis informed by institutional theory. *International Perspectives on Education and Society*, 7, 335-366.
- Christensen, GS. (2004). *What Matters for Immigrant Achievement Cross-Nationally? A Comparative Approach Examining Immigrant and Non-Immigrant Student Achievement* (Doctoral dissertation). California: Stanford University, United States.
- Conchas, GQ. (2001). Structuring failure and success: Understanding the variability in Latino school engagement. *Harvard Educational Review*, 71(3), 475-504.
- Crul, M, & Holdaway, J. (2009). Children of immigrants in schools in New York and Amsterdam: The factors shaping attainment. *Teachers College Record*, 111(6), 1476-1507.
- DerSimonian, R, & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7, 177-188.
- Driessen, G, & Dekkers, H. (1997). Educational opportunities in the Netherlands: Policy, student's performance and issues. *International Review of Education*, 43(4), 299-315.
- Dronkers, J, & Levels, M. (2007). Do School Segregation and School Resources Explain Region-of-Origin Differences in the Mathematics Achievement of Immigrant Students? *Educational Research and Evaluation*, 13(5), 435-462.

- Dronkers, J, Levels, M, & de Heus, M. (2014). Migrant pupils' scientific performance: the influence of educational system features of origin and destination countries. *Large-scale Assessments in Education*, 2(3), 1–28.
- Fox, J, & Weisberg, S. (2011). *An R Companion to Applied Regression* (2nd ed.). Thousand Oaks, CA: Sage. Available from <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>.
- Hanushek, EA, & Kimko, DD. (2000). Schooling, labor-force quality, and the growth of nations. *American Economic Review*, 90, 1184–1208.
- Hedges, LV. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107–128.
- Hedges, LV. (1992). Meta-analysis. *Journal of Educational Statistics*, 17(4), 279–296.
- Hedges, LV, & Vevea, JL. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, 3(4), 486–504.
- Heus, M, Dronkers, J, & Levels, M. (2009). *Immigrant pupils' scientific performance-the influence of educational system features of countries of origin and destination*. San Domenico di Fiesole, Italy: European University Institute.
- Higgins, J, Thompson, SG, Deeks, JJ, & Altman, DG. (2003). Measuring inconsistency in Meta-analysis. *British Medical Journal*, 327, 557–560.
- Hutchison, G, & Schagen, I. (2007). Comparisons between PISA and TIMSS – Are We the Man with Two Watches? In T Loveless (Ed), *Lessons Learned – What International Assessments Tell Us about Math Achievement*. The Brookings Institution: Washington, DC.
- IDB Analyzer (Version 2). (2009). *Computer Software and Manual*. Hamburg, Germany: International Association for the Evaluation of Educational Achievement.
- IEA. (2001a). *Progress in International Reading Literacy Study – PIRLS [Data file]*. Retrieved from [http://timss.bc.edu/pirls2001i/PIRLS2001\\_Pubs\\_UG.html](http://timss.bc.edu/pirls2001i/PIRLS2001_Pubs_UG.html).
- IEA. (2001b). *Trends in International Mathematics and Science Study – TIMSS [Data file]*. Retrieved from <http://timss.bc.edu/timss1995.html>.
- Lee, SJ. (2002). Learning “America”: Hmong American high school students. *Education and Urban Society*, 34(2), 233–246.
- Levels, M, & Dronkers, J. (2008). Educational performance of native and immigrant children from various countries of origin. *Ethnic and Racial Studies*, 31(8), 1404–1425.
- Levels, M, Dronkers, J, & Kraaykamp, G. (2008). Immigrant Children's Educational Achievement in Western Countries: Origin, Destination, and Community Effects on Mathematical Performance. *American Sociological Review*, 73(5), 835–853.
- Ma, X. (2003). Measuring up: Academic performance of Canadian immigrant children in reading, mathematics, and science. *Journal of International Migration and Integration*, 4(4), 541–576.
- Marks, G. (2005). Accounting for immigrant non-immigrant differences in reading and mathematics in twenty countries. *Ethnic and Racial Studies*, 28(5), 925–946.
- Martin, S. (1999). *Economic Integration of Immigrants: A North American-European Comparison* (Discussion Paper prepared for Transatlantic Learning Community). Washington, DC: Migration Workgroup.
- Martin, MO, & Kelly, DL (Eds.). (1996). *TIMSS 1995 Technical Report*. Chestnut Hill, MA: Boston College.
- Martin, MO, Mullis, IVS, & Kennedy, AM (Eds.). (2003). *PIRLS 2001 Technical Report*. Chestnut Hill, MA: Boston College.
- Minoiu, N, & Entorf, H. (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.
- OECD. (2003a). *PISA 2003 Technical Report*. Paris: OECD.
- OECD. (2006a). *Where immigrant students succeed: A comparative review of performance and engagement in PISA 2003*. Paris: OECD.
- OECD. (2006b). *Programme for International Student Assessment – PISA [Data file]*. Retrieved from [http://www.oecd.org/pages/0,3417,en\\_32252351\\_32236130\\_1\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/pages/0,3417,en_32252351_32236130_1_1_1_1_1,00.html).
- OECD. (2006c). *PISA 2006 Technical Report*. Paris: OECD.
- OECD. (2010a). *Closing the gap for immigrant students: Policies, practice, and performance*. Paris: OECD.
- OECD. (2010b). *International Migration Outlook – SOPEMI 2010*. Paris: OECD.
- OECD PISA Website. [<http://www.oecd.org/pisa/aboutpisa>]
- PIRLS & TIMSS Website. [<http://pirls.bc.edu/index.html>]
- Portes, A. (1997). Immigration theory for a new century: Some problems and opportunities. *International Migration Review*, 31(4), 799–825.
- Portes, A, & MacLeod, D. (1996). Educational progress of children of immigrants: The roles of class, ethnicity, and school context. *Sociology of Education*, 69(4), 255–275.
- Portes, A, & MacLeod, D. (1999). Educating the second generation: Determinants of academic achievement among children of immigrants in the United States. *Journal of Ethnic and Migration Studies*, 25(3), 373–396.
- R Core Team. (2014). *R: A language and environment for statistical computing (version 3.1.0)*. Vienna, Austria: R Foundation for Statistical Computing. Available from <http://www.R-project.org>.
- R Development Core Team. (2011). *R: A language and environment for statistical computing (version 2.14.1 Patched)*. Vienna, Austria: R Foundation for Statistical Computing. Available from <http://www.R-project.org>.
- Rangvid, BS. (2007). Sources of immigrants' underachievement: Results from PISA—Copenhagen. *Education Economics*, 15(3), 293.
- Rangvid, BS. (2010). Source country differences in test score gaps: Evidence from Denmark. *Education Economics*, 18(3), 269–295.
- Raudenbush, SW. (2009). Analyzing effect sizes: random-effects models. In HM Cooper, LV Hedges, & JC Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 295–315). New York: Russell Sage.
- Rumbaut, RG. (2004). Ages, life stages, and generational cohorts: Decomposing the immigrant first and second generations in the United States. *International Migration Review*, 38(3), 1160–1205.
- Schneeweis, N. (2006). *On the integration of immigrant children in education* (Paper presented at the First Summer School of the Marie Curie Research Training Network 'The Economics of Education and Education Policy in Europe'. Padova, Italy 16-18 June 2006).

- Schnepf, SV. (2007). Immigrants' educational disadvantage: an examination across ten countries and three surveys. *Journal of Population Economics*, 20(3), 527–545.
- Shadish, WR, & Haddock, CK. (2009). Combining estimates of effect sizes. In H Cooper, LV Hedges, & JC Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 257–277). New York, NY: Russell Sage.
- Thompson, CG, Reta Sánchez, A, Becker, BJ, & Lang, LB. (2011). *A Meta-analysis of the immigrant achievement gap: An analysis of PISA, TIMSS, and PIRLS* (Paper presented at the annual meeting of the Comparative and International Education Society, Montreal, Canada).
- United Nations Development Programme (UNDP). (2009). *Human development report 2009: Overcoming barriers: Human mobility and development*. NY: United Nations.
- Viechtbauer, W. (2010a). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48.
- Viechtbauer, W. (2010b). *metafor (version 1.6-0) [R]*. Available from <http://cran.r-project.org/web/packages/metafor/index.html>.
- Warren, JR. (1996). Educational inequality among White and Mexican-origin adolescents in the American southwest: 1990. *Sociology of Education*, 69(2), 142–158.
- Wößmann, L. (2003). Schooling Resources, Educational Institutions and Student Performance: the International Evidence. *Oxford Bulletin of Economics & Statistics*, 65(2), 117–170.
- Zinovyeva, N, Felgueroso, F, & Vazquez, P. (2008). *Immigration and students' achievement in Spain* (Fedea Report). Fundación de Estudios de Economía Aplicada.

doi:10.1186/s40536-014-0007-2

**Cite this article as:** Andon et al.: A quantitative synthesis of the immigrant achievement gap across OECD countries. *Large-scale Assessments in Education* 2014 2:7.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](http://springeropen.com)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.