The Impact of National and School Contextual Factors on the Academic Performance of Immigrant Students

Holmes Finch [1]*, Maria E. Hernández Finch [2] and Brooke Avery [2]

[1]Educational Psychology, Ball State University, Muncie, IN, United States, [2]Ball State University, Muncie, IN, United States

The issue of immigration has become central to the politics of nations across the world, impacting many aspects of life over the last decade. Researches investigating educational achievement through a cross-national lens have found that immigrant children tend to exhibit lower academic achievement than their native born peers, and that these differences are exacerbated by both family level variables (e.g., socioeconomic status) as well as the school climate. The goal of the current study was to build on earlier work in this area by investigating the nature and degree to which national attitudes towards immigration have changed over time, and whether any such changes were associated with academic achievement for immigrant and native born students. In particular, the relationship between changing attitudes towards immigration and the achievement gap between native and immigrant students. Results of the study demonstrated that nations with more negative attitudes towards refugees in general, and those for which these attitudes became more negative over time had greater achievement gaps than did those nations with more positive attitudes. In addition, these change trajectories moderated relationships between teacher attitudes towards multiculturalism and academic achievement.

Keywords: PISA, immigration, achievement gap, fragile states index, multicultualism

INTRODUCTION

Immigration has become an increasingly key issue across many nations during the second decade of the 21st century. The movement of individuals from the Middle East and Africa has been associated with political backlash in Europe, the United States, Australia among other regions (The Economist, 2019). Of course, such movements of people are not new to the last decade, as many nations have seen consistent immigration for many years, with some attendant issues associated with acculturation by both the new residents and those who were born within the host countries. A key component in the progress of individuals in the United States, both native born and immigrants, is access to and the ability to take advantage of quality educational opportunities (Pivovarova and Powers, 2019). As is discussed in more detail below, much research has examined the impact of immigration status on the academic performance of immigrants to various nations across the world.

In order to assess the impacts of immigration on both the nation to which individuals move and the nation that they left, a variety of tools have been developed. One of these is a component in the fragile states index (FSI), which is developed and maintained by the Fund for Peace (FFP). The FSI (The Fund for Peace, 2020) measures the overall fragility of nations across the world using a variety of metrics, one of which assesses the status of immigrants. The primary goal of the current study was to

ascertain the extent to which a nation's treatment of immigrants (as reflected in the refugees and internally displaced persons index score) was associated with the academic performance of immigrant children, after both individual and school level contextual effects (e.g., socioeconomic status) were taken into account. Following is a discussion of prior research into the relationship of immigration status and academic performance. The goals of the current study are then outlined, followed by a description of the study methods, the results of the study, and a discussion of these results are presented.

IMMIGRATION STATUS AND ACADEMIC PERFORMANCE

There have been a number of studies examining the academic performance of children from immigrant families, particularly in comparison with their native born peers. This work has shown that generally speaking, immigrant students had lower mean reading, math, and science achievement test scores across 34 Organization for Economic Co-operation and Development (OECD) nations, including Australia, Belgium, Canada, France, Germany, Japan, Mexico, Norway, Portugal, Spain, Sweden, the United Kingdom, and the United States, among others (e.g., Schleicher, 2006; Levels and Dronkers, 2008; Marx and Stanat, 2011; Shapira, 2012; Andon, et al., 2014; Pivovarova and Powers, 2019; Borgonovi and Ferrara, 2020; He and Fischer, 2020). This achievement gap has been identified consistently across a number of nations that have quite different immigrant populations. In addition, thsee gaps were found to hold across academic disciplines, and appear to have been most marked for first generation (born outside the host country) students, when compared to those who were second generation (born in the host country, but whose parents were immigrants).

Researchers have identified a number of potential factors that were associated with the achievement test score gap for immigrant students. For example, in research involving more than 20 nations, including Argentina, Switzerland, Costa Rica, Hong Kong, Slovenia, and Turkey, a number of individual/family level variables were associated with lower achievement test performance for immigrant students, including lower income (Andon, et al., 2014; Giannelli and Rapallini, 2016; Radišić et al., 2021). Likewise, in Germany, Austria, the United Kingdom, Sweden, Canada, and the Czech Republic, a lack of facility in the host nation language (Marks, 2005; Schnepf, 2006; Pivovarova and Powers, 2019). Cultural barriers in Canada, Korea, Finland, Greece, Chile, Estonia, the Netherlands, Spain, and Ireland (Rindermann and Thompson, 2014; Bilican Demir and Yildirim, 2020) along with lower parental education attainment (Schnepf, 2006; Andon, et al., 2014) were associated with lower relative educational test performance on the part of immigrants vis-à-vis native born students.

In addition to the impact of these individual level factors, researchers have also found that school effects are also associated with the academic achievement of immigrant students. For example, Martin, et al. (2012) found that in multiple OECD

nations (e.g., Germany, France, Korea, Japan, Russia, the United Kingdom, the United States) the availability of resources for teachers working with immigrant children was associated with the academic performance of their students. Other researchers have found that having a positive learning environment for immigrant students, and positive student attitudes towards their schools were also associated with higher achievement (Schleicher, 2006). Because the Schleicher study did not use an experimental design, it is not possible to determine whether more positive learning environments were causal with respect to higher academic achievement, but the link between the two was found to be relatively large. Rodríguez et al. (2020) examined the extent to which a sense of belonging school was associated with academic achievement for immigrant students in 17 OECD nations, including Hong Kong, Macao, Switzerland, New Zealand, Luxembourg, and the Netherlands. They found that students who felt a sense of welcome and belongingness at their school exhibited higher levels of academic achievement than did those who felt less welcome. When immigrant students were segregated/concentrated in a few schools, the achievement gap vis-à-vis native born students was greater in Portugal, France, Germany, and Canada (Melkonian, et al., 2019; Pivovarova and Powers, 2019).

Researchers have also demonstrated that broader societal attitudes towards immigration, as well as specific government policies were associated with the academic achievement gap between immigrant and native born students. More specifically, immigrant students living in nations where the citizens had more positive attitudes towards immigration had higher levels of academic performance than did immigrants living in nations with less positive attitudes (Rindermann and Thompson, 2014). Likewise, research in a variety of nations including Qatar, Italy, the United Kingdom, the United States, Hungary, Russia, and Serbia, found that students living in nations with more restrictive immigration policies generally performed worse than those living in nations with less restrictive policies (Radišić et al., 2021). Radisic, et al. also demonstrated that nations emphasizing more traditional pedagogical practices exhibited a greater immigrant to native student achievement test performance gap than did those which used a wider array of teaching methods.

STUDY GOALS

There were two primary goals for this study. First, it was of interest to ascertain whether there were different trajectories between 2007 and 2018 in the refugees and internally displaced persons component score across nations for which the fragile states index (FSI) and Programme for International Student Assessment (PISA) were collected. Specifically, the first research question being addressed focused on how these refugee impact scores changed over time, and whether there were subgroups with distinct change trajectories over time. The second goal of this study was to investigate the extent to which membership in subgroups based on change trajectories of the refugee/displaced person index were associated with

student academic performance. In addition, we investigated whether subgroup membership moderated the relationships between selected student and school level variables and academic performance.

Multiple research hypotheses were assessed in this study. First, it was hypothesized that there were different change trajectories in the refugee impact scores over time, with some nations exhibiting increasingly negative attitudes during the time period under study, others exhibiting increasingly positive attitudes, and some nations exhibiting no change over time. In addition, it was hypothesized that the academic performance of immigrant students residing in nations with increasingly negative refugee impacts would be lower than the academic performance of native born students living in these nations. Third, it was hypothesized that the growth trajectory subgroup variable would moderate the relationship between teachers' attitudes toward different cultures and academic achievement. Specifically, we hypothesized that the relationship between teacher attitudes toward different cultures and academic achievement would be positive, but that this relationship would be weaker in nations with negative trajectories in the attitudes towards immigrants. Finally, it was hypothesized that growth trajectory membership would moderate the relationship between the attitudes toward immigration in 2018 and academic achievement in 2018.

METHODS

This project examined data collected through the PISA (OECD, 2018), which is the largest international assessments of students' achievement in mathematics, reading, and science. The data collected from PISA includes demographic information as well as teacher, administrator, and self-reported data. In addition, national level data from the FSI was also used in the study, and is described below. Following is a description of the study participants, the variables used in the analyses, and the analyses used to address the study goals outlined above.

Participants

The sample included a total of 612,004 15-year old students (49.6% female) from the 80 nations participating in PISA (OECD, 2018). More than 20,000 different schools were included in the sample. School samples were selected to be representative within each country. Within each participant nation, all schools registered with the government (both private and public) were included in the sampling frame with the probability of being randomly selected weighted by the number of 15 year old students enrolled, and schools stratified by region of the country. Within each strata, schools were then randomly selected to participate in the PISA project. For each of the selected schools, students were randomly selected from a list of all enrolled 15-year olds. Student samples within each school were chosen using *KeyQuest* software. PISA student data is weighted to be representative of their school and their country, and these weights were used in the current study.

Independent Variables

The independent variables used to address the study goals outlined above included measures of family socioeconomic status (SES), native born status (native), and teacher attitudes towards multiculturalism and equity (SCMCEG) from the PISA 2018 dataset. In addition, the growth trajectory subgroup with respect to treatment of refugees, and the score from this index in 2018 were also included as independent variables in the statistical analyses. These variables are described in more detail below.

PISA Variables

For Family SES, the PISA SES index was derived by OECD (2018) from a factor analysis of variables that include parent education and occupation, home background, as well as possessions in the home (mean = 0, standard deviation = 1). The variable is expressed on a standard normal scale (mean = 0, standard deviation = 1), with higher scores indicating a higher family SES. Native born status was measured by an item on the PISA survey that asked whether students were first generation immigrants, second generation, or more than second generation. The native variable used in this analysis was initially recoded as either more than second generation (1) or first/second generation (0). This coding decision was made in order to differentiate families where both the parents and the children were born in the country of residence (native) from those where at least one member of the family (parents and/or children) were immigrants to the country of residence. As described in the results section, follow up analyses disaggregated this native variable into the original three groups collected by PISA. The SCMCEG variable was based on a scale developed by Hachfeld, et al. (2011) and assessed school leaders' opinions regarding teachers' attitudes towards multiculturalism and equity. The four likert-type items (4 response options) that comprise this scale appear in the **Appendix Table A1**. Higher scores on the SCMCEG score indicated more positive attitudes towards multiculturalism and equity.

Fragile States Index

The attitudes towards refugees and internally displaced persons (RIDP) score from the fragile states index (The Fund for Peace, 2020) was used to assess the pressure on nations that resulted from the inflow and outflow of refugees, as well as movement within nations of internally displaced persons. This variable was designed to measure violence against refugees, sufficiency of resources available to refugees, safety of refugees, and impact of refugee immigration on resource availability for those already residing in the country. The score of each nation included in the study for each year between 2006 and 2018 was used as the variable of interest. Higher RIDP scores (i.e., larger numbers) indicate more negative outcomes (e.g., greater violence, fewer resources, less safety) for refugees. The fragile states index can be accessed at https://fragilestatesindex.org/.

Academics Achievement

Achievement scores for reading and math was used from the individual level PISA database. PISA is not tied to any particular

TABLE 1 | Model fit statistics for GMM solutions.

Model	AIC	BIC	aBIC	BLRT *p*-value
1 class	7341.84	7386.38	7342.04	0 < 0.001
2 classes	7139.38	7209.38	7139.71	0.003
3 classes	7112.09	7194.81	7112.47	0.03
4 classes	7040.33	7135.79	7040.78	0.43
5 classes	7060.55	7161.41	7061.41	0.70

TABLE 2 | Model parameter estimates of the GMM by latent class.

Term	Coefficient	Standard error
Class 1		
Intercept	7.12[a]	0.39
Linear slope	0.01	0.04
Class 2		
Intercept	5.58[a]	0.70
Linear slope	−0.18[a]	0.05
Class 3		
Intercept	2.88[a]	0.91
Linear slope	0.71[a]	0.11
Class 4		
Intercept	2.74[a]	0.81
Linear slope	−0.05	0.09

[a]Statistically significant, α = 0.05.

curriculum, but is designed to examine students' higher level thinking skills, such as analysis, synthesis, and evaluation, through the application of knowledge and skills to real-life situations. The 2-h test contains a mixture of multiple-choice and open-ended items. PISA is methodologically complex and state-of-the-art in development (OECD, 2018). PISA proficiency scales were created using Item Response Theory. Research work (Kankaraš and Moors, 2014) has demonstrated that the quality of measurement of the achievement variables was essentially equivalent across nations, which is crucial for the current study, given its emphasis on estimating relationships among variables. For this reason, it is possible to have confidence in the cross-national meaning of the achievement test scores, given that we are not comparing their means, but rather examining relationships among them.

Data Analysis

In order to determine whether there were distinct trajectories in the RIPD scores, a series of growth mixture models (GMM) were fit to the data using a maximum likelihood estimator. Models with from 1 to 5 classes were considered, and for each the AIC, BIC, and sample size adjusted BIC (aBIC) information indices were calculated. In addition, the bootstrap likelihood ratio test (BLRT) test was used to compare the fit of adjacent models (e.g., 1 vs. 2 class, 2 vs. 3 classes, etc.). The null hypothesis being tested by the BLRT was that the statistical fit of the models yielded the same fit to the data. Thus, a rejection of the null would indicate that the fit of the two models to the data differed, and the information indices were used to determine which number of classes yielded the best fit to the data. With regard to the information indices, the model with the lowest value was deemed to provide the best fit to the data after applying a penalty for model complexity. Once the optimal model was identified, the resulting classes were retained for use in the multilevel model described below. The GMM was fit using Mplus, version 8 (Muthén and Muthén, 2020).

Because students were nested within multiple systems, including schools and countries, a 3-level multilevel regression model with level-1 being students, level-2 being schools, and level-3 being countries was fit to the data in order to ascertain the nature of relationships among the independent variables and reading achievement score. It should be noted that the same model was fit to the data treating the mathematics achievement score as the dependent variable, and the results were nearly identical to those for the reading test. Therefore, only the results for the reading test are reported here. The independent variables described above served as the fixed effects variables in the analysis, including national growth trajectory group, national 2018 RIPD score, school SCMCEG score, student native born

status, and family SES. A random intercepts model was fit to the data with the random effects being the variances associated with school and nation. Prior to fitting the full model including all of the independent variables, a null model including only the variances associated with nation, school, and random error was first fit to the data. The null model provided an estimate of the intraclass correlation (ICC) for reading test scores. The proportion of variance explained in the outcome variables (R^2), and the proportion of reduction in variance compared to the null model (ICC_Δ; Roberts et al., 2011) were calculated for the full model. These data analyses were conducted using SPSS, version 27 (IBM, 2020), and sampling weights were used.

The assumption of normality of the errors was assessed using a QQ-plot and was found to hold, as was homogeneity of variance which was checked using a plot of the residual and predicted values obtained from the model. Given that these assumptions were met, the use of maximum likelihood estimation was deemed to be appropriate. Collinearity among the predictor variables was assessed using the variance inflation factor (VIF), with values in excess of 10 being indicative of collinearity (Fox, 2016). None of the VIF values for this sample exceeded 2, meaning that collinearity was not present. Across the sample, 1.6% of the data were missing. In keeping with recommendations in the literature (e.g., Snijders and Bosker, 2012), missing data was dealt with using full information maximum likelihood (FIML) to obtain model parameter estimates. Grand mean centering was used, per standard practice (Heck and Thomas, 2015).

RESULTS

Hypothesis 1: Differing Change Trajectories in Refugee Impact Scores

The results of the GMM revealed that the 4 class solution yielded the best fit to the data, based on the information indices and the results of the LRT (**Table 1**). More specifically, the 4 class solution had the lowest AIC, BIC, and aBIC values. In addition, this was the first model in the sequence from 1-5 that did not have a statistically significant BLRT result. Recall that the null hypothesis for this test is that there is not a difference in the model fit to the data for the current model (e.g., 4 classes) and the

TABLE 3 | Random effects estimates for null and full models.

Random effect	Estimate	Standard error	95% confidence interval
Null model (school ICC = 0.35, country ICC = 0.17)			
Error	8329.75	18.13	8294.29, 8365.36
School	6062.15	74.14	5918.57, 6209.22
Nation	2839.46	455.54	2073.36, 3888.63
Full Model (School ICC = 0.32, Country ICC = 0.10, Fixed effects R^2 = 0.29)			
Error	7132.23	18.53	7095.99, 7168.65
School	3871.02	58.04	3758.93, 3986.54
Nation	1208.73	235.16	825.52, 1769.81

TABLE 4 | Hypothesis test results for fixed effects in the full model.

Term	DF	F	p
Intercept	1	413.73	<0.001
Non-native	1	1737.68	<0.001
SES	1	16959.34	<0.001
SCMCEG (Teacher multiculturalism/equity)	1	189.08	<0.001
FSI immigration 2018	1	0.425	0.52
GMM class	3	3.96	0.01
Non-native × FSI immigration 2018	1	0.08	0.78
Non-native × Teacher equity	1	0.01	0.94
GMM class × Teacher equity	3	33.81	<0.001
GMM class × FSI immigration 2018	3	1.03	0.27
Non-native × GMM class	3	101.91	<0.001
Non-native × GMM class × Teacher equity	6	1.60	0.14

model with one additional class (e.g., 5 classes). The statistically significant result for the 3 class solution indicates that 4 classes fit the data better than 3 classes, whereas the non-significant 4 class solution means that adding a fifth class does not improve the fit. Taking the information indices and hypothesis test results together, further discussion will be focused on this solution.

Table 2 includes the model intercept and coefficient for linear growth for each latent class from the GMM. The intercept is an estimate of the mean RIDP in the first year of analysis (2006), with the coefficient reflecting the mean change from 1 year to the next. Class 1 was characterized by having the largest mean starting value in the immigrants score (least favorable treatment of refugees in 2006) with no change over time. The nations in class 2 were characterized by having the second largest mean RIDP in 2006, and their score declined (improved) over time. On average the scores for the nations in this class declined by 0.18 each year. The nations in class 3 had the second lowest mean starting score (exhibited the second best treatment of refugees on average in 2006), but had an increase in scores of 0.71 each year on average (i.e., treatment of refugees became less positive over time). Finally, class 4 had the lowest mean RIDP in 2006 and exhibited no change in the scores over time. Latent class membership by nation appears in **Appendix Table A2**. Taken together, these results support the first hypothesis that there would be different change trajectories in the refugee impact scores over time.

Multilevel Modeling of Achievement Data

As described in the Methods section, a multilevel model was used in order to investigate the relationship between the GMM latent

TABLE 5 | Mean reading score and standard error by GMM class and non-native status.

| Non-native status	GMM class	Mean	Standard error
Non-native	1	396.77	21.12
	2	399.23	10.16
	3	444.84	15.12
	4	449.48	14.37
Native	1	436.59	21.09
	2	425.51	10.09
	3	483.60	15.06
	4	453.05	14.35

Cohen's d values for mean comparisons between native and non-native examinees. GMM class 1, Native > Non-native, Cohen's d = 0.41. GMM class 2, Native > Non-native, Cohen's d = 0.26. GMM class 3, Native > Non-native, Cohen's d = 0.39. GMM class 4, Native > Non-native, Cohen's d = 0.03.

classes and reading achievement, while controlling for individual and school level variables. **Table 3** includes the random effects estimates for the null model for which there were no predictor variables, and the full model for which all of the predictors described in the Methods section were included. The variances associated with school and country were statistically significant for both models (the confidence intervals did not include 0), indicating that there were differences in mean reading scores across schools within nations, and across the nations themselves. The ICC values in **Table 3** reflect the proportion of variance in the reading test scores that were associated with both school and nation. When no fixed effects were included in the model, school accounted for 35% of the reading score variance and nation accounted for 17%. When the full fixed effects portion of the model was specified, school accounted for 32% of the reading score variance and nation accounted for 10%. The set of fixed effects variables accounted for approximately 29% of the variance in the reading test scores.

Hypothesis 2: Academic Achievement of Immigrant Students Residing in Nations With Negative Attitudes Towards Immigrants Will Perform Worse Than Native Born Students in Those Nations

Table 4 includes the hypothesis test results for the fixed effects included in the full multilevel model with PISA reading score serving as the dependent variable. The main effects of non-native status, family SES, school mean teacher equity score (SCMCEG), GMM class, as well as the interactions between GMM class and SCMCEG, and GMM class and non-native status were statistically significantly associated with the reading test score. The only statistically significant main effect that was also not involved in a statistically significant interaction was family SES, which had a coefficient of 30.45 (**Table 5**), indicating that examinees from families with a higher SES also had higher scores on the reading test.

The mean reading scores by non-native status and country GMM class appear in **Table 5** and **Figure 1**. The means between the native and non-native examinees within each latent class were compared using simple contrasts with a Bonferroni correction
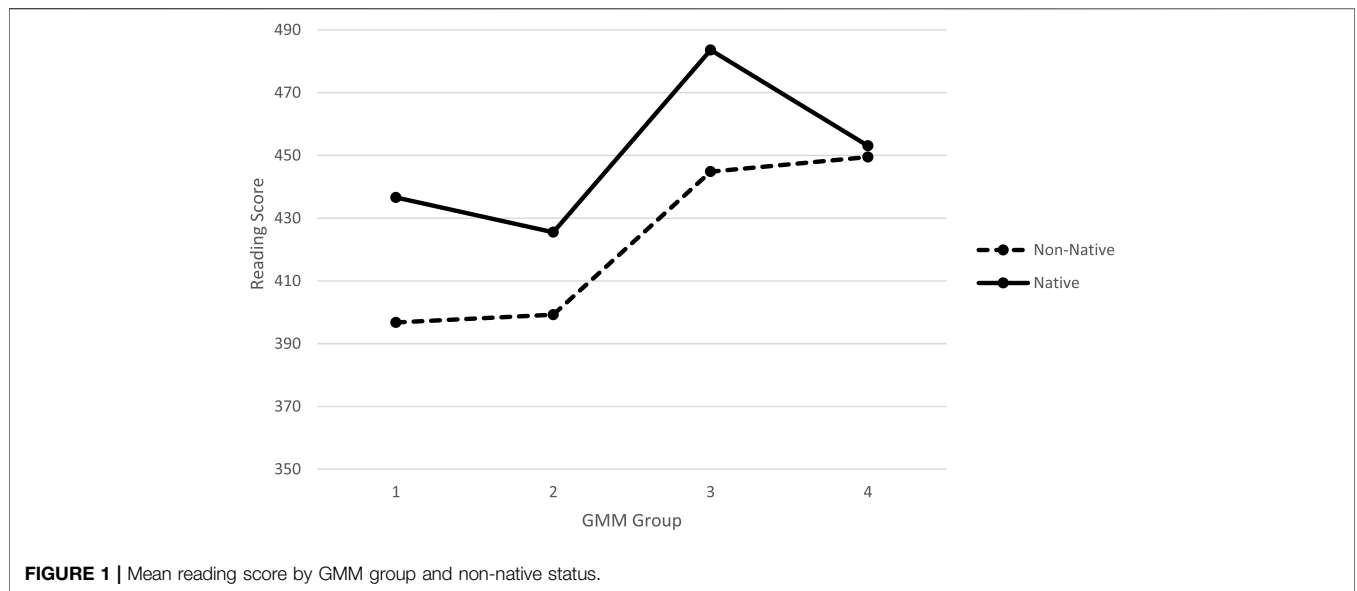
FIGURE 1 | Mean reading score by GMM group and non-native status.

used to control the Type I error rate. The results of these tests revealed that the mean reading scores for native born examinees were significantly higher than those of immigrant students within GMM classes 1, 2, and 3, but not for class 4. The values of Cohen's *d* for the comparison of group means within latent class appear as footnotes in **Table 5**. Based on Cohen's (1988) guidelines for interpreting these values, the difference in mean scores for classes 1–3 fell within the small range, whereas the mean difference for class 4 was negligible in size.

In order to more fully explore the relationships between GMM latent class, native/non-native status, and reading test scores, the multilevel model was refit to the data treating non-native status in its original form as collected by PISA with 3 categories, including first generation, second generation, and native. The results for this model were very similar to those for the model including the dichotomous native/non-native variable, and will therefore not be discussed in more detail here. A set of contrasts comparing the means of the three non-native status categories within GMM latent classes revealed that within each class, the means for the native category were significantly higher than those of second generation students. In turn, the mean reading scores of the second generation students were significantly higher than those for first generation students within each class. The means for the paired comparisons within GMM latent class appear in **Table 6** and **Figure 2**, and the Cohen's *d* values for these comparisons appear in **Table 7**. In addition, the means for the native status groups by GMM latent class appear in **Figure 2**. For latent classes 1 and 3, the Cohen's *d* values for the native vs first generation comparisons fell within Cohen's (1988) large guidelines. Cohen's *d* values for the other mean comparisons in latent classes 1–3 were in the small range, including native vs second generation and second generation vs first generation. In contrast, the Cohen's *d* values were in the negligible range for GMM class 4. These results support the second hypothesis that immigrant students living in nations with more negative

TABLE 6 | Mean reading score and standard error by GMM class and residency status.

Non-native status	GMM class	Mean	Standard error
1st Generation	1	378.03	20.98
	2	378.67	10.59
	3	419.66	15.33
	4	445.96	14.80
2nd Generation	1	398.14	21.77
	2	401.83	10.07
	3	448.28	15.31
	4	453.05	14.08
Native	1	436.59	21.09
	2	425.51	10.09
	3	483.60	15.06
	4	455.25	14.35

Cohen's d *values for mean comparisons between native and non-native examinees.*

attitudes towards immigration will perform worse than native born students living in those nations.

Hypothesis 3: The Relationship Between Teacher Attitudes Toward Different Cultures and Academic Achievement Will Be Moderated by National Attitudes Toward Immigration

The coefficients for the teacher equity score by GMM latent class appear in **Table 8**. Across latent classes, students attending schools where teachers had a more positive attitude towards cultural diversity and equity (higher SCMCEG scores) also had higher reading test scores. These coefficients were compared between latent classes using *z*-tests with the type I error rate controlled using the Bonferroni procedure. These results revealed that the relationship between teacher equity and the reading test score for class 3 (5.74) was significantly smaller than for classes 1 (6.61), 2 (6.84), and 4 (7.96). In addition, the coefficient for class 4
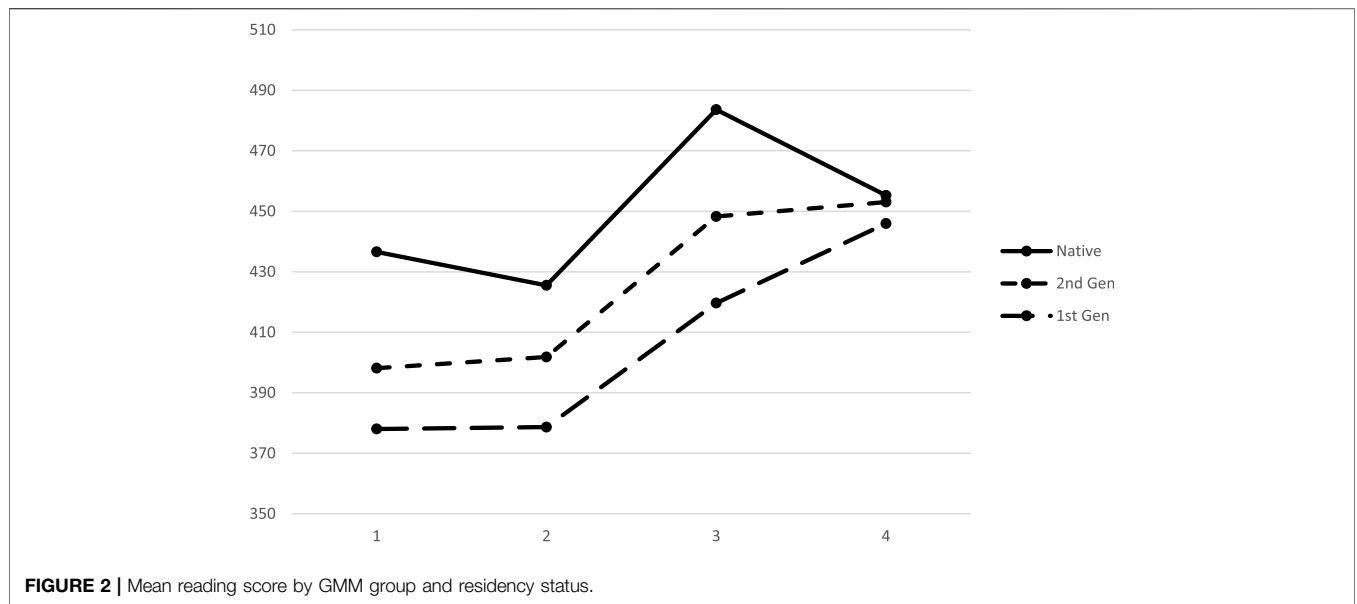
FIGURE 2 | Mean reading score by GMM group and residency status.

TABLE 7 | Mean contrast results and Cohen's *d* effect sizes for non-native status groups.

GMM class	Comparison 1	d	Comparison 2	d	Comparison 3	d
Class 1	Native v 2nd gen	0.36	Native v 1st gen	0.55	2nd gen v 1st gen	0.19
Class 2	Native v 2nd gen	0.25	Native v 1st gen	0.45	2nd gen v 1st gen	0.22
Class 3	Native v 2nd gen	0.37	Native v 1st gen	0.64	2nd gen v 1st gen	0.29
Class 4	Native v 2nd gen	0.02	Native v 1st gen	0.09	2nd gen v 1st gen	0.06

TABLE 8 | Covariate coefficients, standard errors, and 95% confidence intervals for highest level statistically significant effects.

Term	Estimate[a]	Standard error	95% confidence interval
Intercept	478.63[a]	8.92	460.75, 496.51
SES	30.45[a]	0.16	30.13, 30.77
SCMCEG: GMM class 1	6.61[a]	0.33	5.96, 7.26
SCMCEG: GMM class 2	6.84[a]	0.32	6.22, 7.45
SCMCEG: GMM class 3	5.74[a]	0.53	4.70, 6.79
SCMCEG: GMM class 4	7.96[a]	0.28	7.40, 8.52

[a]*Statistically significant, α = 0.05.*

was significantly larger than those for classes 1 and 2, which did not differ from one another. Thus, the relationship between teacher attitudes towards multiculturalism and equity were more strongly positively related to reading test score for those in class 4 than in the other classes, whereas the relationship was weakest for those in class 3. This result supports hypothesis 3.

Hypothesis 4: The Growth Trajectory Membership Will Moderate the Relationship Between Attitudes Towards Immigration and Academic Achievement

The fourth hypothesis was that the change in attitudes toward immigration over time would moderate the relationship between attitudes in 2018 and academic achievement in that year. The

statistically non-significant interaction between GMM class and FSI 2018 immigration score (**Table 4**) indicates that such moderation was not found to be present in the current study. Therefore, hypothesis 4 was not supported in this study.

DISCUSSION

The results of the analyses presented above revealed the existence of four distinct trajectories in attitudes and treatment of refugees across nations participating in the PISA testing program. These groups differed both in terms of the mean starting RIDP score, as well as the change in these scores over time. Two of the groups were characterized by relatively higher RIDP values (worse treatment of refugees), with nations in one experiencing

improvements over time (Latent class 2) and the other having no change (Latent class 1). The other two latent classes had lower mean starting scores (relatively better treatment of refugees), with one having no mean change in the scores (Latent class 4) and nations in the other having their RIDP increase in value (increasingly worse treatment of refugees between 2006 and 2018; Latent class 3). These results support the first hypothesis that was assessed by this study, namely that there are different subgroups in the population of nations with respect to how attitudes and treatment of refugees evolved over time.

The second research hypothesis investigated by this work focused on whether for nations where attitudes towards refugees became worse over time, non-native born students would exhibit worse academic performance than native students. This hypothesis was partially supported by the results presented above. It was true that the reading means for non-native examinees were significantly lower than for native born examinees in latent class 3, which experienced a significant decline in the treatment of refugees. However, such differences were also found for latent classes 1 and 2, both of which had high scores but either no change, or improvement in RIDP scores. Indeed, only for latent class 4 was there not a statistically significant difference in reading score means between the native and non-native groups. Taken together, these results suggest that for nations with relatively positive treatment/ attitudes regarding refugees and for which this did not change between 2006 and 2018 there was no difference in reading achievement test performance for native and non-native examinees, but for all other nations native born individuals tended to perform better on the reading assessment. This finding supports results from earlier studies that demonstrated a relationship between societal and school treatment of immigrants and their performance on academic tasks (Marks, 2005; Rindermann and Thompson, 2014; Radišić et al., 2021).

The third hypothesis, that the latent class would moderate the relationship between teacher attitudes towards multiculturalism and equity (SCMCEG) and academic achievement was supported. The relationship between SCMCEG and reading achievement was most strongly positive for examinees living in latent class 4, and weakest for those in latent class 3. Thus, student attendance at schools where teachers valued multiculturalism more was associated with better reading achievement, and this relationship was strongest in nations with more positive and stable treatment/attitudes towards refugees. Interestingly, the weakest relationship between reading scores and SCMCEG was found for the nations that exhibited the best treatment of refugees in 2006, but which saw the greatest diminution of this support over time. Those nations that exhibited the most negative treatment of refugees in 2006, but that either improved or remained the same in this regard also had a stronger positive relationship between teacher attitudes towards multiculturalism and equity than in nations with more positive refugee treatment, but where that treatment degraded over time. Prior research (Schleicher, 2006; Martin, et al., 2012; Melkonian, et al., 2019; Pivovarova and Powers, 2019; Rodríguez et al., 2020) has shown that school level factors such as a positive learning environment and resource availability are associated with the academic performance of immigrant students. The current study furthers this work by showing that these school level effects on immigrant's academic performance is itself influenced by the broader cultural attitude toward and treatment of immigrants.

Finally, the fourth hypothesis, that latent class would moderate the relationship between national attitudes toward refugees in 2018 (the year of the PISA test) and reading achievement was not supported. Indeed, there was not a relationship between the single year RIDP score and reading achievement across the latent classes, in addition to there not being a statistically significant interaction between latent class and the 2018 RIDP score.

The results of this study both support and amplify earlier findings that have been reported in the literature. Prior work has demonstrated that there is an achievement gap between immigrant and native born students (e.g., Andon, et al., 2014; Borgonovi and Ferrara, 2020). This study also found evidence of such an achievement gap for many nations, though not all. Specifically, these results suggest that in nations that exhibit generally positive treatment of refugees consistently over multiple years, the gap does not exist. In contrast, in nations where this treatment has traditionally been less positive, or where it began positive and then deteriorated over time, the native versus immigrant achievement gap was present. As noted above, these results supported earlier work showing that a positive school climate (in the form of teacher attitude toward multiculturalism in the current study) was associated with greater academic achievement (Schleicher, 2006; Rodríguez et al., 2020). However, this effect was lower for schools in countries with more negative and/or declining attitudes towards refugees. Finally, it should be noted that these results regarding the relationship of national level measures of refugee treatment and academic achievement were found to be present above and beyond family SES, which has consistently been related to academic achievement, as was the case here.

Limitations and Directions for Future Research

The current study was designed to ascertain whether there existed distinct subgroups within the population of nations with respect to the course of change in treatment of refugees and displaced persons over time, and whether these groupings were associated with academic achievement. The results presented above serve, we believe, to extend on earlier work investigating the achievement gap between immigrant and native born students. As with all research efforts, the current study has limitations that future research should be conducted to address. First, although this study was designed to ascertain the role of national level treatment of refugees on academic achievement, it would be of interest to learn the extent to which these policies impact individual schools. Future research should be conducted to investigate how national policies impact school practice, and in turn how these practices specifically impact student achievement. Future research should also examine whether the results presented here also carry over to other national level testing programs such as TIMMS and PIRLS. In order to

investigate the issue from a more nuanced perspective, results for specific nations could also be studied. This study treated all immigrants within a nation as being monolithic. However, it is clear that this is not truly the case. Therefore, future work should more closely consider immigration and home language status for specific nations to determine whether students from different cultural and language backgrounds exhibit divergent academic achievement for nations in the various latent classes. Finally, it is very possible that there exist relationships between the structures and practice of national educational systems, changes in national attitudes towards immigration, and academic performance of immigrants. Therefore, future research should more carefully examine the relationships among these factors in an attempt to identify what aspects of educational systems are associated with the attitudes towards immigration change trajectories identified by the mixture model.

CONCLUSION

Results of this study highlight the relationship between broad national trends in the treatment of and attitudes toward refugees, and academic achievement for non-native students. Specifically, in nations where the treatment of refugees degraded over time, there was a wider gulf between the reading performance of native and non-native students than was the case for nations where this treatment did not change and was relatively positive. In addition, non-native students also exhibited lower average reading scores than their native peers in nations where the treatment of refugees was relatively negative and stable, or even slightly improved over time. In short, how refugees are treated by the nations in which they reside, particularly if that treatment becomes worse over time, appears to be associated with greater gaps in the academic achievement of native and non-native born students. This trend was even true for those who were born in the country of residence but whose parents were not. Therefore, education policy makers should carefully consider how their nation's policies towards refugees may deleteriously impact first and second generation students, and what mechanisms could be put into place to mitigate these negative effects. In contrast, non-native students living in nations for which treatment of refugees was stable and positive appear to perform comparably to their native born classmates.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.oecd.org/pisa/data/2018database/.

AUTHOR CONTRIBUTIONS

HF ran statistical analysis and did the majority of writing. MH developed core research ideas, identified data sources, and wrote portions of the manuscript. BA did literature search and formatted references.

REFERENCES

Andon, A., Thompson, C. G., and Becker, B. J. (2014). A Quantitative Synthesis of the Immigrant Achievement Gap across OECD Countries. *Large-Scale Assessments Educ.* 2 (1). doi:10.1186/s40536-014-0007-2

Bilican Demir, S., and Yildirim, O. (2020). Indirect Effect of Economic, Social, and Cultural Status on Immigrant Students' Science Performance through Science Dispositions: A Multilevel Analysis. *Educ. Urban Soc.* 53 (3), 336–356. doi:10.1177/0013124520928602

Borgonovi, F., and Ferrara, A. (2020). Academic Achievement and Sense of Belonging Among Non-native-speaking Immigrant Students: The Role of Linguistic Distance. *Learn. Individ. Differ.* 81, 101911. doi:10.1016/j.lindif.2020.101911

Fox, J. (2016). *Applied Regression Analysis and Generalized Linear Models.* Thousand Oaks, CA: Sage.

Giannelli, G. C., and Rapallini, C. (2016). Immigrant Student Performance in Math: Does it Matter where You Come from? *Econ. Educ. Rev.* 52, 291–304. doi:10.1016/j.econedurev.2016.03.006

Hachfeld, A., Hahn, A., Schroeder, S., Anders, Y., Stanat, P., and Kunter, M. (2011). Assessing Teachers' Multicultural and Egalitarian Beliefs: The Teacher Cultural Beliefs Scale. *Teach. Teach. Educ.* 27 (6), 986–996. doi:10.1016/j.tate.2011.04.006

He, J., and Fischer, J. (2020). Differential Associations of School Practices with Achievement and Sense of Belonging of Immigrant and Non-immigrant Students. *J. Appl. Dev. Psychol.* 66, 101089. doi:10.1016/j.appdev.2019.101089

Heck, R. H., and Thomas, S. L. (2015). *An Introduction to Multilevel Modeling Techniques: MLM and SEM Approaches Using Mplus.* 3rd ed. Boca Raton, FL: Routledge/Taylor & Francis Group.

IBM (2020). *IBM SPSS Statistics for Windows.* Version 27.0. Armonk, NY: IBM Corp.

Kankaraš, M., and Moors, G. (2014). Analysis of Cross-Cultural Comparability of PISA 2009 Scores. *J. Cross-Cultural Psychol.* 45 (3), 381–399. doi:10.1177/0022022113511297

Levels, M., and Dronkers, J. (2008). Educational Performance of Native and Immigrant Children from Various Countries of Origin. *Ethnic Racial Stud.* 31 (8), 1404–1425. doi:10.1080/01419870701682238

Marks, G. N. (2005). Accounting for Immigrant Non-immigrant Differences in reading and Mathematics in Twenty Countries. *Ethnic Racial Stud.* 28 (5), 925–946. doi:10.1080/01419870500158943

Martin, A. J., Liem, G. A. D., Mok, M. M. C., and Xu, J. (2012). Problem Solving and Immigrant Student Mathematics and Science Achievement: Multination Findings from the Programme for International Student Assessment (PISA). *J. Educ. Psychol.* 104 (4), 1054–1073. doi:10.1037/a0029152

Marx, A. E., and Stanat, P. (2011). Reading Comprehension of Immigrant Students in Germany: Research Evidence on Determinants and Target Points for Intervention. *Read. Writ.* 25 (8), 1929–1945. doi:10.1007/s11145-011-9307-x

Melkonian, M., Areepattamannil, S., Menano, L., and Fildago, P. (2019). Examining Acculturation Orientations and Perceived Cultural Distance Among Immigrant Adolescents in Portugal: Links to Performance in reading, Mathematics, and Science. *Soc. Psychol. Educ.* 22 (4), 969–989. doi:10.1007/s11218-019-09506-5

Muthén, L. K., and Muthén, B. O. (2020). *Mplus User's Guide.* Eigth Edition. Los Angeles, CA: Muthén & Muthén.

OECD (2018). *PISA 2018 Results: What Students Know and Can Do.* Available: https://www.oecd.org/pisa/publications/pisa-2018-results-volume-i-5f07c754-en.htm (Accessed September 17, 2021).

Pivovarova, M., and Powers, J. M. (2019). Generational Status, Immigrant Concentration and Academic Achievement: Comparing First and Second-Generation Immigrants with Third- Plus Generation Students. *Large-Scale Assessments Educ.* 7 (1). doi:10.1186/s40536-019-0075-4

Radišić, J., Selleri, P., Carugati, F., and Baucal, A. (2021). Are Students in Italy Really Disinterested in Science? A Person-Centered Approach Using the PISA 2015 Data. *Sci. Educ.* 105 (2), 438–468. doi:10.1002/sce.21611

Rindermann, H., and Thompson, J. (2016). The Cognitive Competences of Immigrant and Native Students across the World: An Analysis of Gaps, Possible Causes and Impact. *J. Biosoc. Sci.* 48 (1), 66–93. doi:10.1017/S0021932014000480

Roberts, J. K., Monaco, J. P., Stovall, H., and Foster, V. (2011). "Explained Variance in Multilevel Models," in *Handbook for Advanced Multilevel Analysis*. Editors J. J. Hox and J. K. Roberts, 219–230.

Rodríguez, S., Valle, A., Gironelli, L. M., Guerrero, E., Regueiro, B., and Estévez, I. (2020). Performance and Well-Being of Native and Immigrant Students. Comparative Analysis Based on Pisa 2018. *J. Adolesc.* 85, 96–105. doi:10.1016/j.adolescence.2020.10.001

Schleicher, A. (2006). Where Immigrant Students Succeed: A Comparative Review of Performance and Engagement in Pisa 20031. *Intercultural Educ.* 17 (5), 507–516. doi:10.1080/14675980601063900

Schnepf, S. V. (2006). Immigrants' Educational Disadvantage: an Examination across Ten Countries and Three Surveys. *J. Popul. Econ.* 20 (3), 527–545. doi:10.1007/s00148-006-0102-y

Shapira, M. (2012). An Exploration of Differences in Mathematics Attainment Among Immigrant Pupils in 18 OECD Countries. *Eur. Educ. Res. J.* 11 (1), 68–95. doi:10.2304/eerj.2012.11.1.68

Snijders, T. A. B., and Bosker, R. J. (2012). *Multilevel Analysis*. Thousand Oaks, CA: Sage Publishers.

The Economist (2019). *Why the Arguments against Are So Popular*. London, UK: Economist.

The Fund for Peace (2020). *Fragile States Index*. Available at: https://fragilestatesindex.org/ (Accessed September 17, 2021).

APPENDIX

TABLE A1 | SCMCEG items.

1. It is important for students to learn that people from other cultures can have different values.
2. Respecting other cultures is something that students should learn as early as possible.
3. In the classroom, it is important that students of different origins recognise the similarities that exist between them.
4. When there are conflicts between students of different origins, they should be encouraged to resolve the argument by finding common ground.

TABLE A2 | Nation by latent class.

Class	Nations
1	Azerbaijan, Bosnia, Colombia, Croatia, Georgia, Israel/West Bank, Jordan, Lebanon, Morocco, Philippines, Serbia, Thailand, Turkey
2	Bulgaria, China, Costa Rica, Dominican Republic, Indonesia, Malaysia, Malta, Mexico, Moldova, Montenegro, Peru, Russia, Saudi Arabia, Vietnam
3	Albania, Austria, Belarus, Brazil, Germany, Greece, Hungary, Italy, Japan, Macedonia, Poland, Romania, Slovenia, Sweden, Ukraine
4	Argentina, Australia, Belgium, Brunei, Canada, Chile, Czech Republic, Denmark, Estonia, Finland, France, Iceland, Ireland, Kazakhstan, Latvia, Lithuania, Luxembourg, Netherlands, New Zealand, Norway, Panama, Portugal, Qatar, Singapore, Slovak Republic, South Korea, Spain, Switzerland, UAE, United Kingdom, United States