Diversity in Schools

Immigrants and the performance of US-born students

David Figlio (Rochester) Paola Giuliano (UCLA) Riccardo Marchingiglio (Analysis Group)
Paola Sapienza (Northwestern) Umut Ozek (RAND)

June 2023

Motivation:

- Never-ending debate in policy and academia on the impact of immigration on natives' welfare.
- The overwhelming majority of the literature focuses on labor market impact on native adults.
- We focus on a different effect: The exposure to immigrant peers on native public school students: arguably a first-order impact.
- In the US alone 23 percent of the student population comes from an immigrant family.
- This work helps to estimate the payoffs from immigration policies.

Contribution

- Papers studying the effects of foreign-born peers on natives' outcomes in school:
  - European context: negative [Jensen and Rasmussen (2011), Brunello and Rocco (2013), Ballatore et al. (2018), Tornello (2016), Bossavie (2020)] or no effect [(Ohinata et al. (2013), Geay et al. (2013) and Schneeweis (2015)].
  - Israel: negative effect of immigrants on native Israeli students' likelihood of passing high school matriculation exam (Gould, Lavy and Paserman, (2009)).
  - United States: negative relationship between natives' test scores and immigrant share at the school level (Schwartz and Stiefel (2001)), but positive effect on the high school completion of natives (Hunt, (2016)).

Identification challenge: Endogenous sorting of both **immigrants** and **natives**. Much of previous literature addresses **immigrant** sorting but (so far) not **native** sorting.

Contribution

- Our main contribution: We make use of matched birth and school records in Florida to take into account, for the first time, sorting of both **immigrants** and **natives**.
    - Birth records allow us to compare siblings within the same family.
    - We exploit within-family variation and plausibly exogenous school-to-school transitions.
    - This complements previous work that has been able to account for **endogenous sorting of immigrants** but not yet endogenous sorting of natives.
    - We also differentiate between being exposed to **different types of immigrants** in the classroom.
- Endogenous sorting of natives is a big deal: **We find evidence that native families experiencing more immigrants in kindergarten move their children to another school**.

This paper in a nutshell

- Data
    - administrative
    - longitudinal
    - family identifiers
- Identification strategy exploits:
    - Sibling comparison
    - Holding fixed time-varying family characteristics (as well as time-varying school and grade characteristics), compare different cumulative exposures to first generation immigrants
    - Instrumental variable approach: use aggregate school-to-school transition probabilities to build predicted exposures for each kid at each subsequent grade, starting from the first at which she is first observed
    - Two siblings will therefore have the same transition matrix but a different exposure to immigrants, which depends on the specific cohort they are in
- Results:
    - Positive relationship (larger in math), mainly driven by disadvantaged groups
    - Immigrants do not negatively affect the achievement of US born students even when their academic achievement is lower than their US born classmates.

This paper in a nutshell

- Data
    - administrative
    - longitudinal
    - family identifiers

- Identification strategy exploits:
    - Sibling comparison
    - Holding fixed time-varying family characteristics (as well as time-varying school and grade characteristics), compare different cumulative exposures to first generation immigrants
    - Instrumental variable approach: use aggregate school-to-school transition probabilities to build predicted exposures for each kid at each subsequent grade, starting from the first at which she is first observed
    - Two siblings will therefore have the same transition matrix but a different exposure to immigrants, which depends on the specific cohort they are in

- Results:
    - Positive relationship (larger in math), mainly driven by disadvantaged groups
    - Immigrants do not negatively affect the achievement of US born students even when their academic achievement is lower than their US born classmates.

This paper in a nutshell

- Data
  - administrative
  - longitudinal
  - family identifiers
- Identification strategy exploits:
  - Sibling comparison
  - Holding fixed time-varying family characteristics (as well as time-varying school and grade characteristics), compare different cumulative exposures to first generation immigrants
  - Instrumental variable approach: use aggregate school-to-school transition probabilities to build predicted exposures for each kid at each subsequent grade, starting from the first at which she is first observed
  - Two siblings will therefore have the same transition matrix but a different exposure to immigrants, which depends on the specific cohort they are in
- Results:
  - Positive relationship (larger in math), mainly driven by disadvantaged groups
  - Immigrants do not negatively affect the achievement of US born students even when their academic achievement is lower than their US born classmates.

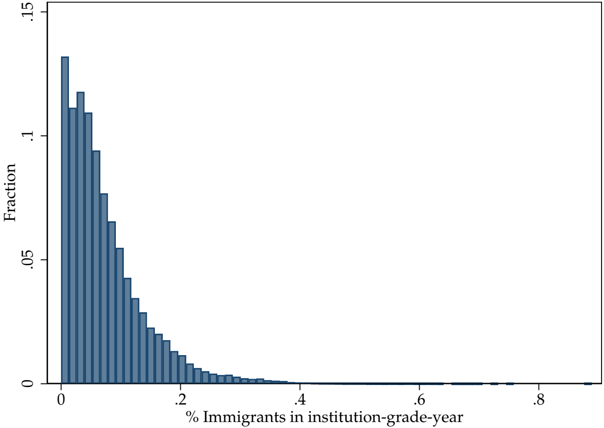
Data

- Individual-level administrative data from the Florida Department of Education Data Warehouse from the academic year 2002-2003 through the academic year 2011-2012:
  - K-12 students who attended FLPS born between 1994-2002
  - longitudinal data
- Matched birth records for those born in Florida (using SSN, names, DOB)
- Florida has the fourth highest number of immigrants in the United States, and Florida's immigrant population is more diverse than most places.
- Outcome of interest: Florida Comprehensive Assessment Test (FCAT) in mathematics and reading:
  - Standardized, with mean 0 and standard deviation 1, at the grade-year level over the entire population of students
  - Regressions in levels from grade 3 to grade 10

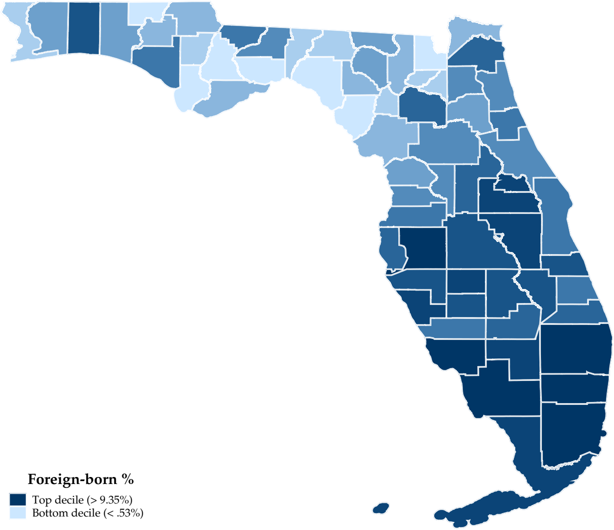
Definition of immigrants and natives

- Immigrants: children born outside the United States
- Natives: children born in the U.S. and speaking English at home (note: We've looked at other variations on this theme too)
- Treatment of Puerto Rican-born students is not obvious: They are US citizens but are also culturally distinct from many other US citizens, and nature of their selection to schools, school selection of other citizens in response, and effects on peers might all be different from other US citizens.
  - Therefore, we explore the consequences of treating Puerto Ricans as "immigrants" vs. "natives".
  - In practice, results are extremely similar regardless of treatment of Puerto Ricans.

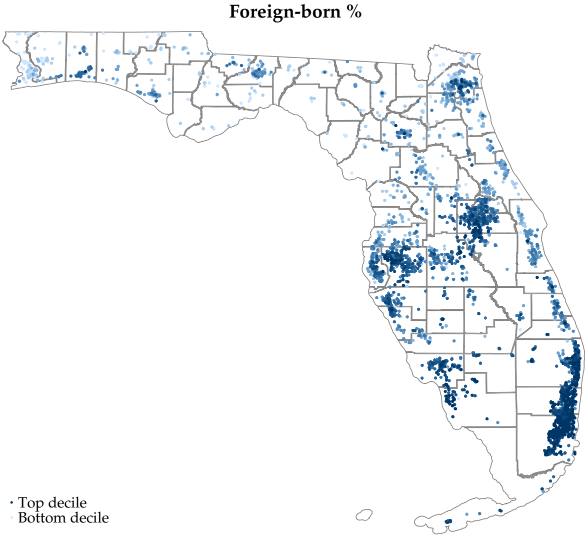
Exposure of US-born students to foreign-born peers



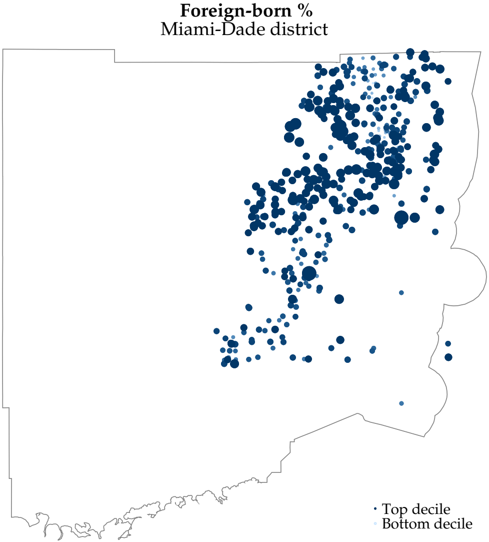
Distribution of foreign born students by district



Distribution of foreign-born students: within district



Distribution of foreign-born students: within district



Countries of origin

	Overall	Nat. White Majority*	Nat. Hisp. Majority	Nat. Black Majority
	Top 10 Immigrants' countries of origin			
1.	Cuba (16%)	Mexico (13%)	Cuba (45%)	Haiti (41%)
2.	Mexico (10%)	Puerto Rico (7%)	Colombia (9%)	Jamaica (13%)
3.	Haiti (10%)	Colombia (7%)	Mexico (7%)	Mexico (6%)
4.	Colombia (8%)	Germany (5%)	Venezuela (6%)	Puerto Rico (4%)
5.	Puerto Rico (6%)	Cuba (4%)	Puerto Rico (4%)	Cuba (3%)
6.	Venezuela (5%)	Canada (4%)	Honduras (3%)	Honduras (3%)
7.	Jamaica (3%)	Haiti (3%)	Dominican Rep. (3%)	Dominican Rep. (2%)
8.	Peru (3%)	Venezuela (3%)	Argentina (3%)	Bahamas (2%)
9.	Argentina (2%)	Brazil (3%)	Peru (3%)	Colombia (2%)
10.	honduras (2%)	China (3%)	Nicaragua (2%)	Japan (1%)
Top-10 Cumul.	65%	51%	86%	78%

* Native white majority indicates that only school-specific cohorts with more than 50% white U.S.-born are selected. The third and fourth column are analogously constructed.

Selection

Ethnic groups

	Overall	Nat. White Majority	Nat. Hisp. Majority	Nat. Black Majority
		Top 3 Immigrants' ethnic groups		
1.	Hispanic (61%)	Hispanic (45%)	Hispanic (92%)	Black (64%)
2.	Black (16%)	White (30%)	Black (3%)	Hispanic (27%)
3.	White (13%)	Asian (13%)	White (3%)	Asian (5%)
Top-3 Cumul.	90%	88%	98%	96%

Cumulative exposure

What is the impact of being exposed to a larger share of immigrants during a student's school career?

Right-hand-side variable:

$$E_{isgt} = \sum_{g' \leq g} IMMIGRANTSHARE_{isg't} * e^{(1-(\lambda*(g-g')))} / \sum_{g'g} e^{(1-(\lambda*(g-g')))}$$

Left-hand-side: Standardized test scores in mathematics and reading (Y_{istg}).

A cumulative exposure measure has the advantages of

- smoothing out abrupt changes in class composition
- accounting for lagged effects

Existing literature does not provide direction on the specific size of λ.
We investigate the full range of values, today present case of $\lambda = 0$.

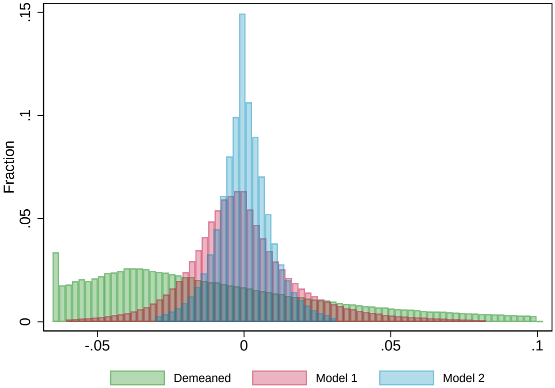
Main specification

$$Y_{istg} = \alpha_{st} + \alpha_{gt} + \theta_{f(i),t} + \beta \boldsymbol{E}_{istg} + \boldsymbol{\delta}' \boldsymbol{W}_{istg} + \varepsilon_{istg} \tag{1}$$

- school by year FEs
- grade by year FEs
- family by year FEs
- $\boldsymbol{W}_{istg}$ individual and family controls (e.g., gender, age in months, birth order, free lunch, race)

The regressions are run on a subset of observations such that there are at least 2 siblings in each family, each year.

Identifying variation: Exposure



Demeaned: $\mathbb{P}(X - \bar{X})$
Model 1: $\mathbb{P}(X|\text{institution} \times \text{year, year} \times \text{grade})$.
Model 2: $\mathbb{P}(X|\text{institution} \times \text{year, year} \times \text{grade, family} \times \text{year})$.

Estimates: Math

	Math standardized scores (3-10 grades)				
Foreign-born exp.	-0.125**	0.018	0.076*	0.289***	0.224***
	(0.053)	(0.042)	(0.040)	(0.054)	(0.074)
Individual Controls	Y	Y	Y	Y	Y
School × Year FEs	Y	Y	Y	Y	Y
Grade × Year FEs	Y	Y	Y	Y	Y
Race FEs	N	Y	Y		
Lunch Status	N	Y	Y		
Mother's Educ. FEs	N	N	Y		
Family FE				Y	
Family × Year FE					Y
Observations	1,347,286	1,347,286	1,344,541	1,347,286	1,347,286
R^2	0.302	0.359	0.379	0.682	0.769
Mean RHS	0.060	0.060	0.060	0.060	0.060
SD RHS	0.052	0.052	0.052	0.052	0.052
β	-0.006	0.001	0.004	0.015	0.012

Individual controls include gender, age in months, special education, birth order FEs. Standard errors are clustered at the cohort-school level. Partial persistence: decay

Estimates: Reading

	Math standardized scores (3-10 grades)				
Foreign-born exp.	-0.194**	-0.026	0.041	0.174***	0.108***
	(0.049)	(0.039)	(0.037)	(0.048)	(0.064)
Individual Controls	Y	Y	Y	Y	Y
School × Year FEs	Y	Y	Y	Y	Y
Grade × Year FEs	Y	Y	Y	Y	Y
Race FEs	N	Y	Y		
Lunch Status	N	Y	Y		
Mother's Educ. FEs	N	N	Y		
Family FE				Y	
Family × Year FE					Y
Observations	1,450,138	1,450,138	1,447,278	1,450,138	1,450,138
R^2	0.303	0.356	0.377	0.667	0.752
Mean RHS	0.061	0.061	0.061	0.061	0.061
SD RHS	0.053	0.053	0.053	0.053	0.053
β	-0.010	-0.001	0.002	0.009	0.006

Individual controls include gender, age in months, special education, birth order FEs. Standard errors are clustered at the cohort-school level. Partial persistence: decay

Does high "immigrant exposure" really mean "segregation"?

	(1) weighted	(2) High-Seg now	(3) Low-Seg now	(4) High-Seg first	(5) Low-Seg first
Foreign-born cumul. exp.		0.282*	0.358***	0.312*	0.299**
		(0.169)	(0.112)	(0.173)	(0.125)
Foreign-born cumul. exp. (weighted)	0.235**				
	(0.097)				
Beta coefficient	0.009	0.009	0.022	0.012	0.019
Individual Controls	Y	Y	Y	Y	Y
School × Year FEs	Y	Y	Y	Y	Y
Grade × Year FEs	Y	Y	Y	Y	Y
Family × Year FE	Y	Y	Y	Y	Y
Observations	1,450,139	1,450,139	1,447,279	1,450,139	1,450,139
Observations	1,450,139	1,450,139	1,447,279	1,450,139	1,450,139
R-squared	0.761	0.768	0.781	0.777	0.786
Dependent Variable (mean)	0.034	0.034	0.034	0.034	0.034
Dependent Variable (sd)	0.992	0.992	0.992	0.992	0.992
RHS (mean)	0.0367	0.0381	0.0889	0.0477	0.0770
RHS (sd)	0.0373	0.0311	0.0618	0.0372	0.0639

More evidence on selection and sorting

We expect selection of natives into schools based on immigrant exposure (especially given what we know about post-kindergarten sorting.)

We know from the first table that selection is likely negative: low achieving native students are associated with larger shares of immigrants.

But, what sub-populations are responsible for the sorting? Let's split the sample by ethnicity and socioeconomic status.

Splitting the sample by race

	Math standardized scores (3-10 grades)				
	Black sub-population				
Foreign-born exp.	0.511***	0.495***	0.475***	0.441***	0.385***
	(0.067)	(0.066)	(0.065)	(0.097)	(0.137)
N	399,586	399,586	398,269	399,586	399,586
R^2	0.266	0.273	0.283	0.593	0.716
	White sub-population				
Foreign-born exp.	-0.610***	-0.395***	-0.261***	0.209**	0.128
	(0.064)	(0.061)	(0.058)	(0.075)	(0.107)
N	811,790	811,790	810,559	811,790	811,790
R^2	0.263	0.284	0.312	0.671	0.764
Individual Controls	Y	Y	Y	Y	Y
School × Year FEs	Y	Y	Y	Y	Y
Grade × Year FEs	Y	Y	Y	Y	Y
Lunch Status	N	Y	Y		
Mother's Educ. FEs	N	N	Y		
Family FE				Y	
Family × Year FE					Y

Splitting the sample by socio-economic status

	Math standardized scores (3-10 grades)				
	Free or reduced-price lunch sub-population				
Foreign-born exp.	0.367***	0.281***	0.300***	0.445***	0.387***
	(0.053)	(0.050)	(0.049)	(0.074)	(0.102)
N	735,589	735,589	733,624	735,589	735,589
R^2	0.250	0.280	0.293	0.620	0.728
	Full-price lunch sub-population				
Foreign-born exp.	-0.462***	-0.426***	-0.298***	-0.003	-0.035
	(0.067)	(0.065)	(0.061)	(0.080)	(0.113)
N	611,698	611,698	610,918	611,698	611,698
R^2	0.218	0.235	0.270	0.672	0.763
Individual Controls	Y	Y	Y	Y	Y
School $\times$ Year FEs	Y	Y	Y	Y	Y
Grade $\times$ Year FEs	Y	Y	Y	Y	Y
Lunch Status	N	Y	Y		
Mother's Educ. FEs	N	N	Y		
Family FE				Y	
Family $\times$ Year FE					Y

Additional selection issues

Family fixed effects solves the selection issue if family makes the school decision for all children in the same way.

Families may make differential school choice decisions based on the characteristics of each single child.

- If families send the highest achieving child to a school with fewer immigrants, the estimated coefficient on the share of immigrants would be downward biased.
- If families have egalitarian preferences (Becker and Tomes (1976)) and send the lowest achieving child to a school with fewer immigrants, the estimated coefficient on the share of immigrants would be upward biased.

Instrumental variable strategy

Intuition for predicted exposure:

1. Fix the initial school
2. Build aggregate school-to-school transition matrices
3. Predict exposures at each subsequent grade starting from the first observed
4. Compare siblings who started in the same school (in possibly different years/grades)
5. Two siblings will therefore have the same transition matrix bit a different exposure to immigrants, which depends on the specific cohort they are in

Predicted Exposure: Construction

- For each pair of consecutive grades g and $g+1$, π_{kj} is the probability that a student in school k at grade g ends up in school j at grade $g+1$.
- For each grade g and time t, $\boldsymbol{W}(g,t)$ is a vector of school characteristics.
- N_s is the total number of schools in the sample.

Transition matrix from grade g to grade $g+1$

$$\mathbb{P}(g+1|g) = \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} & \dots & \pi_{1N_s} \\ \pi_{21} & x_{22} & \pi_{23} & \dots & \pi_{2N_s} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \pi_{N_s1} & \pi_{N_s2} & \pi_{N_s3} & \dots & \pi_{N_sN_s} \end{bmatrix}$$

$$\boldsymbol{W}(g,t) = \begin{bmatrix} w_1(g,t) & w_2(g,t) & w_3(g,t) & \dots & w_{N_s}(g,t) \end{bmatrix}'$$

Predicted Exposure: Construction

Relevant objects:

$$\left\{ \underset{(N_s \times N_s)}{\mathbb{P}(g+1|g)} \right\}_{g=0}^{11} \qquad \text{12 } (N_s \times N_s)\text{-transition matrices}$$

$$\left\{ \left\{ \underset{(N_s \times 1)}{\boldsymbol{W}(g,t)} \right\}_{g=0}^{12} \right\}_{t=2002}^{2011} \qquad \text{130 } (N_s \times 1)\text{-vectors}$$

Building the predicted exposure at $(\tilde{g}, \tilde{t})$ based on Markov chains for given (g_0, t_0):

$$\underset{(N_s \times 1)}{\boldsymbol{Z}(\tilde{g}, \tilde{t})} = \mathbb{E}\left[\boldsymbol{W}(\tilde{g}, \tilde{t}) | (g_0, t_0)\right] = \underset{(N_s \times N_s)}{\left(\prod_{g=g_0}^{\overset{\curvearrowright}{\tilde{g}-1}} \mathbb{P}(g+1|g) \right)} \underset{(N_s \times 1)}{\boldsymbol{W}(\tilde{g}, \tilde{t})}$$

IV Estimates

	RF	OLS	IV
		Math	
Foreign-born exposure	0.139***	0.336***	0.320***
(predicted for RF)	(0.067)	(0.068)	(0.155)
N	821,892	821,892	821,892
R^2	0.668	0.668	-
Individual Controls	Y	Y	Y
Family $\times$ Initial School	Y	Y	Y
Family $\times$ Grade	Y	Y	Y

Individual controls include gender, age in months, special education. Standard errors are clustered at the cohort-initial-school level.

Heterogeneity: Relative standing and absolute performance

	US-born speaking English	Immigrants who go to school with them
	Average math scores	
Whole sample	0.050	0.006
White US-born	0.305	0.093
Black US-born	-0.495	-0.180
Full-price lunch US born	0.475	0.170
Free or reduced-price lunch US-born	-0.303	-0.137

Heterogeneity by cross-country differences in immigrant math performance

	Math standardized scores (3-10 grades)				
Restriction:	Full sample	No free lunch	Free lunch	White	Black
Foreign-born exposure	0.214***	-0.047	0.371***	0.132	0.391***
	(0.078)	(0.118)	(0.108)	(0.112)	(0.144)
Immigrant performance index	0.037***	0.031**	0.037***	0.032***	0.036**
	(0.008)	(0.013)	(0.011)	(0.011)	(0.017)
Observations	1,271,257	585,025	686,232	764,912	374,370
R^2	0.778	0.770	0.740	0.774	0.730

Heterogeneity by cross-country differences in immigrant misbehavior

	Math standardized scores (3-10 grades)				
Restriction:	Full sample	No free lunch	Free lunch	White	Black
Foreign-born exposure	0.210***	-0.048	0.365***	0.131	0.384***
	(0.078)	(0.118)	(0.108)	(0.112)	(0.144)
Immigrant misbehavior index	-0.253***	-0.213**	-0.283***	-0.204**	-0.257**
	(0.069)	(0.108)	(0.092)	(0.092)	(0.128)
Observations	1,271,257	585,025	686,232	764,912	374,370
R^2	0.778	0.770	0.740	0.774	0.730

Heterogeneity by cross-country differences in immigrant long-term orientation (Figlio et al, 2019)

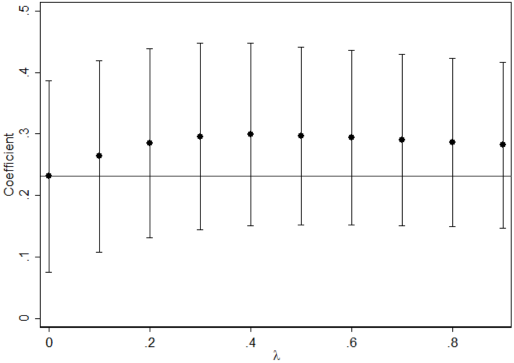
	Math standardized scores (3-10 grades)				
Restriction:	Full sample	No free lunch	Free lunch	White	Black
Foreign-born exposure (LTO above US)	0.632***	0.272	0.941***	0.491**	1.023***
	(0.167)	(0.220)	(0.264)	(0.209)	(0.370)
Foreign-born exposure (LTO below US)	0.201	-0.012	0.292*	0.110	0.423*
	(0.123)	(0.184)	(0.174)	(0.178)	(0.247)
Immigrant performance index	0.028***	0.022	0.028**	0.024**	0.025
	(0.009)	(0.013)	(0.012)	(0.011)	(0.018)
Observations	1,271,257	585,025	686,232	764,912	374,370
R^2	0.778	0.770	0.740	0.774	0.730

Conclusion

- We use within-family variation and a novel identification strategy to identify the impact of foreign-born exposure to native students' outcomes.
- The coefficient is mostly driven by low-SES and African-American students.
- Selection of US-born and immigrants in schools generate interesting patterns of interactions:
- Low SES US-born students mostly interact with higher (than them) performing immigrants
- Absolute performance (academic and behavioral) correlates positively with the performance of all US born students, independently from their SES, but it does not explain the correlation between the presence of immigrants and the performance of US born students (especially low SES)
- Relative performance may explain our heterogeneous results but we are not able to test this hypothesis directly

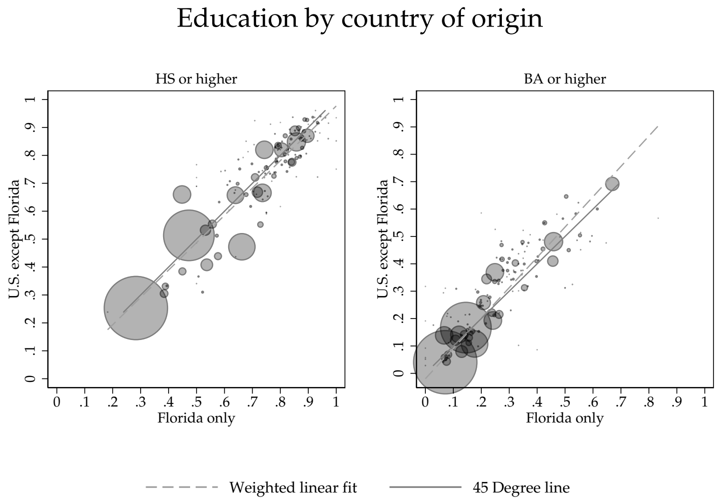
Partial persistence: a model of decay

$$Exposure_G = \frac{\sum_{g=g_{min}}^{G} X_g e^{1-\lambda(G-g)}}{\sum_{g=g_{min}}^{G} e^{1-\lambda(G-g)}}$$



Back

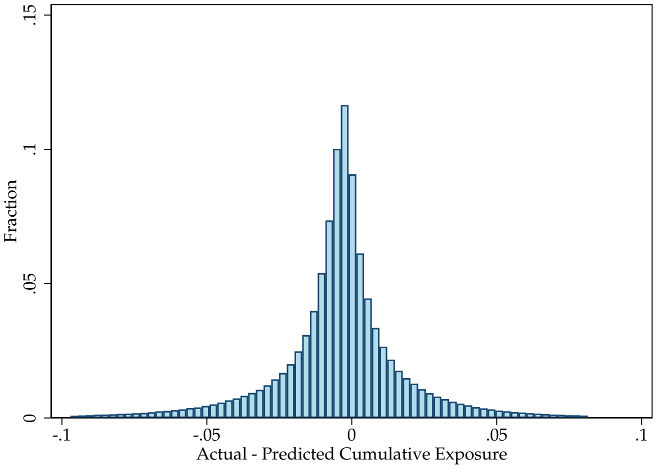
Education selection



Source: U.S. Census 2000, 5% (IPUMS)

Back

Deviations



Back

Natives SES vs foreign-born SES

	Math standardized scores (3-10 grades)					
	Free or Reduced price lunch sub-population					
Foreign-born exp.	0.539***	0.382***	0.360***	0.385***	0.371***	0.501***
	(0.056)	(0.052)	(0.051)	(0.079)	(0.112)	(0.122)
Cumulative share of low-SES among foreign-born peers	-0.271***	-0.178***	-0.133***	-0.033***	-0.035***	-0.005
	(0.009)	(0.008)	(0.008)	(0.009)	(0.012)	(0.013)
N	667,360	667,360	665,613	667,360	667,360	667,360
R^2	0.259	0.288	0.302	0.639	0.744	0.744
	No reduced price sub-population					
Foreign-born exp.	-0.165**	-0.167**	-0.109*	-0.002	0.021	0.096
	(0.068)	(0.066)	(0.063)	(0.084)	(0.120)	(0.127)
Cumulative share of low-SES among foreign-born peers	-0.250***	-0.209***	-0.131***	-0.016*	-0.018	0.003
	(0.010)	(0.009)	(0.009)	(0.009)	(0.013)	(0.013)
N	579,622	579,622	578,897	579,622	579,622	579,622
R^2	0.218	0.234	0.269	0.677	0.771	0.772
Individual contr., S-Y, G-Y	Y	Y	Y	Y	Y	Y
Race FEs	N	Y	Y			
Mother's Educ. FEs	N	N	Y			
Family FE				Y		
Family × Year FE					Y	Y
Exposure controls	N	N	N	N	N	Y

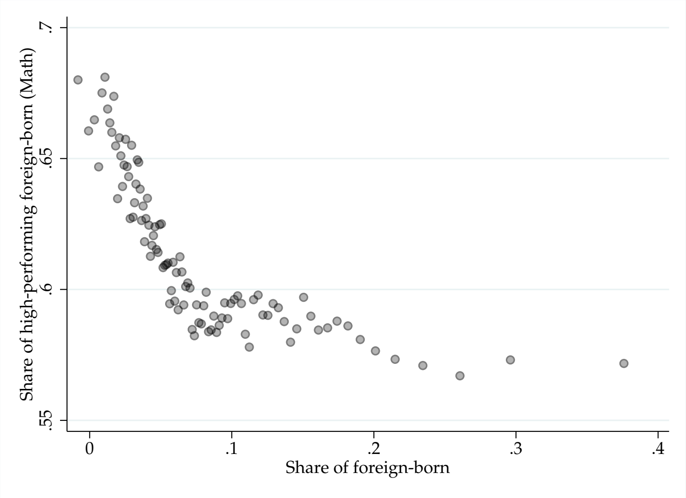
Back

Summary Statistics

	Mean	Median	SD
Free/Reduced price lunch	0.54	–	–
Female	0.50	–	–
Special Education	0.14	–	–
White	0.60	–	–
Black	0.28	–	–
Hispanic	0.07	–	–
Mother's years of schooling	–	12	–
Age in months	138.59	137	25.23
% Black exposure	0.24	0.16	0.24
% Hispanic exposure	0.19	0.14	0.18
% Asian exposure	0.02	0.02	0.02
% LEP exposure	0.05	0.03	0.07
% Free/Red. p. lunch exposure	0.55	0.56	0.24

Back

Does the "quality" of immigrants matter?



Back