Education-Occupation Mismatch and Nativity Inequality Among Highly Educated U.S. Workers

Xiaoguang Li and Yao Lu

Online Appendix

Section A. The selection of the realized match approach

There are three common approaches for determining the matched education credentials for any given occupation: realized match, self-assessment, and job analysis (McGuinness 2006). The *self-assessment* approach is based on workers' subjective assessments about the matched education credentials for their current occupations (Duncan and Hoffman 1981). Such information, however, is often unavailable in social surveys. The *job analysis* approach uses objective evaluations of occupational requirements by professional job analysts (Rumberger 1981). However, up-to-date measurements using this approach are usually unavailable. The *realized match* approach, which is most commonly used, is based on the observed educational distribution of workers in each occupation. The matched education (level or field) for an occupation is derived from what workers in a given occupation have typically attained, such as the mode of that distribution. The mode is preferred over the mean because it is less sensitive to outliers and technological change (Kiker et al. 1997). It is also conceptually superior because it captures the most frequently traveled education-occupation pathways (Kerckhoff 1996). Recently, researchers have pursued a new demand-side approach, which assesses the education credentials that employers seek, as indicated on online job postings (Lu and Li 2021). An occupation is classified as college-level if more than 50% (or another percentage) of job postings for that occupation require a bachelor's degree or higher.

We chose the realized match approach for several reasons. First, this approach provides an objective measure that is readily available and can be applied to any dataset that contains information on educational credentials (level and field of study) and occupations (Ortiz and Kucel 2008). Because this approach is data driven, it can measure mismatch at the level of field of study or occupation in a more detailed way than the job analysis approach, which typically provides information on broad fields and occupational groups. Second, the realized mismatch approach can be used to construct both vertical and horizontal mismatch measures. This is not possible when using the job analysis approach in the US context. The Bureau of Labor Statistics (BLS) provides job analyses of educational requirements (level) for occupations (first available in 2014), but it does not provide information on the matched fields of study for occupations. Third, because standards derived from the realized match approach are more readily available, they can easily be updated to reflect changes in requirements due to technological transformations and other factors. Also, the timing of ACS (for the realized match approach) better aligns with the main data for analysis (SIPP) than the BLS standards.

Previous research that compares the three common approaches show that they tend to yield consistent patterns (Lu and Li 2021; McGuinness 2006). We also conducted similar analyses using other criteria (e.g., BLS and demand-side) and found similar differences in mismatch with respect to nativity.

Section B

Table B1. Descriptive statistics (SIPP 1996-2011)

Variables	Percentage or Mean (SD)
Immigration Status	
Native born	90.5
Immigrants	9.5
Immigration Status by Place of Education	
Native born	90.5
US-educated immigrants	2.0
Foreign-educated immigrants	7.5
Immigration Status by Duration of Residence in the U.S. [a]	
Native born	90.8
Immigrants for 0-5 years	2.0
Immigrants for 6 years or more	7.2
Immigration Status by Quality of Tertiary Education (QTE) in Origin Country [b]	
Native born	93.7
Immigrants from high QTE countries	3.0
Immigrants from low QTE countries	3.3
Immigration Status by STEM Fields	
Native born	90.5
Immigrants with STEM degrees	3.7
Immigrants with non-STEM degrees	5.7
Immigration Status by English Proficiency [c]	
Native born	89.5
Immigrants with proficient English	8.6
Immigrants with less proficient English	1.9
Immigration Status by Licensed Fields	
Native born	90.5
Immigrants in licensed fields	3.4
Immigrants in non-licensed fields	6.1
Race/ethnicity	
Non-Hispanic White	85.0
Non-Hispanic Black	5.8
Hispanic	3.3
Asian	5.9
Female	48.0
Age	40.0
	(8.6)
Married	73.2
Years of Education	16.8
	(1.3)
Field of Study [d]	
Business	22.3
STEM	20.1
Health Sciences	5.4
Social Sciences, History, Psychology, Communication	21.5
Education	14.0
Liberal Arts, Humanities, Architecture	7.7
Medicine, Dentistry, Law	4.3
Others	5.0

Work Experience	20.3
	(9.1)
Job Tenure	8.2
	(7.7)
Hourly Wage (log transformed)	3.3
	(0.6)
Public Sector Employment	25.8
Union Membership	13.5
Occupation	
Management and Professional	74.3
Service	3.8
Clerical and Sales	17.6
Production, Farming, and Construction	4.3
Total Number of Real Occupation Changes	0.3
	(0.6)
Panel	
1996 panel	22.4
2001 panel	21.3
2004 panel	30.2
2008 panel	26.1
Metropolitan Area Residency	84.1
Region	
Northeast	21.1
Midwest	26.4
South	33.9
West	18.7
Number of Observations	106,520
Number of Individuals	13,315

Notes: The end year reported in the table is the last year of the SIPP panel used in the analysis.
a. The sample size is 105,408 because of missing data for duration of residence.
b. Restricted to SIPP 1996-2004 because country of origin information is unavailable in SIPP 2008. The sample size is 77,096.
c. Restricted to SIPP 2004-2008 where English proficiency is available. The sample size is 59,312.
d. We show 1-digit ISCED fields of study in this descriptive table to save space, we but used 2-digit ISCED fields of study in all regression models.

Section C. GMM estimation for nativity differences in mismatch

One common strategy to adjust for endogeneity bias is to estimate longitudinal individual fixed-effects (FE) models. This strategy is not feasible in our study because the key variable of interest, immigration status, is time-invariant and thus cannot be estimated. FE models also are untenable for studying the wage consequences of mismatch because of its high persistence. The FE approach relies on sufficient within-individual variance for model identification, but the key variables in the wage regressions are either time-invariant (immigration status) or exhibit limited changes during the observation window (mismatch). The lack of variation over time limits the statistical power and efficiency of FE models.

Therefore, we used the GMM approach to adjust for the potential of endogeneity bias. The GMM approach effectively adjusts for time-varying and time-invariant unobserved heterogeneity by using instrumental variables. Specifically, we used system GMM models, which combine the original level regression equation and its first-difference equation with an instrumental variable approach by using a generalized method of moments. In these models, the lagged levels and lagged first-differences of the regressors serve as instruments for the endogenous explanatory variables. Previous research shows that in system GMM regressions, the lagged levels are valid instruments for the first-difference equation and the lagged differences are valid instruments for the level regression (Arellano and Bover 1995; Blundell and Bond 2000).

Previous research has demonstrated that the system GMM approach generates consistent and efficient estimates and also exhibits several strengths over conventional two-stage instrumental variable models (Arellano and Bond 1991). Specifically, this approach is more efficient because it utilizes additional lags of variables as instruments. It also overcomes the common challenge of identifying specific instrumental variables. Moreover, this method accommodates multiple endogenous variables in the same model and allows for the estimation of time-invariant variables. The system GMM method is the most suitable option when the sample size is considerably larger than the number of waves in the panel (Roodman 2009; Wooldridge 2010), as is the case in our study.

We used the GMM approach to estimate both the incidence and wage consequences of mismatch for robustness checks. The GMM models passed the F-test for weak instruments, the Hansen test of overidentification restrictions, and the Arellano-Bond test for no second-order serial correlation in the first-differenced residuals. These tests suggest that the GMM models are appropriate for our study. The GMM results are generally consistent with the corresponding random-effects models presented in the main paper.

Table C1. GMM models for the incidence of mismatch by nativity

	Model 1	Model 2	
	Vertical Mismatch	Horizontal Mismatch (base category: Horizontal Match)	
		Overmatch	Undermatch
Immigration status (ref. = native born)			
US-educated immigrant	0.018	-0.015	0.003
	(0.013)	(0.010)	(0.022)
Foreign-educated immigrant	0.044***	-0.018	-0.050*
	(0.013)	(0.013)	(0.023)
Control variables	Yes	Yes	Yes
Arellano-Bond test for AR(1)	-5.680***	-6.770***	-8.53***
P-value for AR(1)	0.000	0.000	0.000
Arellano-Bond test for AR(2)	1.430	1.040	-0.990
P-value for AR(2)	0.154	0.298	0.453
Hansen test of overid	49.210	41.880	61.250
P-value for Hansen test	0.814	0.841	0.395
Number of Observations	93,205	93,205	93,205
Number of individuals	13,315	13,315	13,315

Notes: The results from generalized method-of-moments (GMM) are showed in the tables. The Arellano-Bond test shows that the differences of the residual terms have been autocorrelated for first-differences, but not for second-differences. Hence, the null hypothesis of no autocorrelation in the nuisance terms cannot be rejected, indicating that the GMM is suitable to use. The Hansen statistic rejects the null hypothesis that all instrumental variables are exogenous at the 0.05 level. Therefore, the overall model setup is reasonable, and the instrumental variables are valid.

The control variables (for all models) include gender, race/ethnicity, age, age squared, marital status, years of education, working experience, job tenure, total number of occupational changes, public sector employment, union membership, metropolitan area residency, region, and survey panel. Additionally, in Model 1 that estimates vertical mismatch, we also controlled for horizontal mismatch; in Model 2 that estimates horizontal mismatch, we also controlled for vertical mismatch.

*** $p<0.001$, ** $p<0.01$, * $p<0.05$.

Table C2. GMM models for the wage penalties of mismatch by nativity

	Model 1	Model 2
	Hourly Wage (log transformed)	Hourly Wage (log transformed)
Immigration Status (ref. = native born)		
US-educated immigrant	0.012	0.071
	(0.053)	(0.053)
Foreign-educated immigrant	-0.025	-0.041
	(0.067)	(0.060)
Vertical Mismatch (VM)	-0.134***	0.071
	(0.024)	(0.053)
Interaction		
US-educated immigrant × VM	0.088	
	(0.065)	
Foreign-educated immigrant × VM	-0.239**	
	(0.087)	
Horizontal Overmatch (HO)		-0.026
		(0.024)
Horizontal Undermatch (HU)		-0.027
		(0.040)
Interaction		
US-educated immigrant × HO		-0.079
		(0.083)
Foreign-educated immigrant × HO		0.043
		(0.053)
US-educated immigrant × HU		-0.010
		(0.071)
Foreign-educated immigrant × HU		-0.143*
		(0.060)
Control Variables	Yes	Yes
Number of Observations	92,044	92,044
Number of Individuals	13,172	13,172

Notes: The results from generalized method-of-moments (GMM) are showed in the tables. The control variables (for all models) include gender, race/ethnicity, age, age squared, marital status, years of education, field of study, working experience, job tenure, total number of occupational changes, public sector employment, union membership, metropolitan area residency, region, occupation, and survey panel. Additionally, in Model 1 that estimates vertical mismatch, we also controlled for horizontal mismatch; in Model 2 that estimates horizontal mismatch, we also controlled for vertical mismatch.
*** p<0.001, ** p<0.01, * p<0.05.

Section D

Table D1. Common examples of vertical and horizontal mismatch

	Horizontal Match	*Horizontal Overmatch*	*Horizontal Undermatch*
Vertical Match	Life science → Medical scientists; Computer science → Computer programmers; Teaching → Secondary school teachers; Journalism → Editors and reporters	Life science → Administrators; Architecture → Financial managers; Accounting → Education administrators; Social science → Chief executives and administrators	Engineering → Secondary school teachers; Architecture → Social workers; Accounting → Social workers; Law → Editors and reporters
Vertical Mismatch	Engineering → Industrial machinery operators or repairers; Life science → Registered nurses or medical assistants; Business and administration → Payroll clerks; Accounting → Bookkeeping clerks	Computer science → Construction Managers; [a] Teaching → Science technicians; Humanities → Computer support specialists	Math → Sales workers; Computer science → File clerks; Teaching → Sales workers; Humanities → Waiters and waitresses

[a] It is very uncommon for vertical mismatch and horizontal overmatch to co-occur, especially among STEM majors.

Section E. A sensitivity analysis using data from the National Survey of College Graduates

The assignment of foreign-degree status in SIPP is based on the age of arrival and years of education variables. As a sensitivity analysis, we used the National Survey of College Graduates (NSCG) to measure foreign-degree status. NSCG2010 provides direct information on whether the respondent's highest degree was earned in the United States. Using this information, we distinguished among the native born (74%), US-educated immigrants (15%), and foreign-educated immigrants (11%). The NSCG measured the degree of education-occupation match through self-assessments: respondents were asked "To what extent was your work on your principal job held this week related to your highest degree?". The response categories were "Closely related," "Some related," and "Not related." Whereas this variable allows us to obtain a general sense of nativity differences in mismatch using direct information on place of degree, it is not directly comparable to that which is constructed in SIPP because the two variables are based on different approaches (self-assessment in NSCG and realized match in SIPP). Also, the NSCG employs a self-defined occupational classification system (with a particular focus on STEM occupations) and is therefore difficult to map for commonly used occupational classification schemes such as the Census or the SOC codes. This prevents us from measuring education-occupation mismatch in NSCG using the realized match approach. Moreover, the self-reported measure in NSCG could capture both the vertical and horizontal match status of respondents.

Using this question, we distinguished among matched (or "closely related," 61.7%), moderately mismatched (or "some related," 24%) and mismatched (or "not related," 14.3%). The results are presented in Table E. Regarding the incidence of mismatch, foreign-educated immigrants were significantly more likely than native born to undergo education-occupation mismatch (Model 1). Regarding the wage penalty of mismatch, foreign-educated immigrants suffered a higher wage penalty than similarly mismatched native born Americans (Model 2). However, US-educated immigrants were not significantly different from the native born in both the incidence and wage penalties of mismatch. Taken together, these results are broadly consistent with findings based on place of degree, which was derived from age of arrival and years of education in SIPP.

Table E1. Incidence and wage penalty of mismatch by nativity (NSCG2010)

	Model 1 Mlogit (AME) Education-occupation mismatch (base category: Match)		Model 2 OLS Hourly Wage (log transformed)
	Moderately Mismatched	Mismatched	
Immigration Status			
(ref. = native born)			
US-educated immigrant	0.008	-0.004	0.075***
	(0.016)	(0.016)	(0.021)
Foreign-educated immigrant	-0.031	0.058**	-0.059*
	(0.016)	(0.019)	(0.028)
Education-occupation Mismatch			
(ref. = Match)			
Moderately mismatched			0.075***
			(0.021)
Mismatched			-0.059*
			(0.028)
Interaction			
US-educated immigrant × Moderately mismatched			-0.045
			(0.037)
Foreign-educated immigrant × Mismatched			-0.245***
			(0.049)
US-educated immigrant × Moderately mismatched			-0.068
			(0.056)
Foreign-educated immigrant × Mismatched			-0.279***
			(0.051)
Control Variables.	Yes	Yes	Yes
Number of Individuals	60,511	60,511	60,511

Notes: We included as many control variables used in SIPP as possible. The control variables (for all models) include gender, race/ethnicity, age, age squared, marital status, years of education, field of study, working experience, and region.
*** p<0.001, ** p<0.01, * p<0.05.

Section F. Continuous measures of education-occupation mismatch

Table F1. Degree of mismatch and wage penalty by nativity

	Model 1	Model 2
	Degree of Vertical Mismatch	Degree of Horizontal Mismatch
Panel A. Degree of Mismatch Outcomes (AMEs are shown)		
Immigration Status (ref. = native born)		
US-educated immigrant	0.128*	0.185
	(0.054)	(0.095)
Foreign-educated immigrant	0.265***	0.197***
	(0.034)	(0.059)
Control Variables	Yes	Yes
Panel B. Hourly Wage Outcomes (log transformed)		
Immigration Status (ref. = native born)		
US-educated immigrant	0.112**	0.069*
	(0.037)	(0.035)
Foreign-educated immigrant	-0.086***	-0.083***
	(0.021)	(0.021)
Degree of Vertical Mismatch	-0.063***	
	(0.004)	
Interaction		
US-educated immigrant × Degree of vertical mismatch	0.072**	
	(0.022)	
Foreign-educated immigrant × Degree of vertical mismatch	-0.038**	
	(0.012)	
Degree of Horizontal Mismatch		-0.008***
		(0.002)
Interaction		
US-educated immigrant × Degree of horizontal mismatch		0.014
		(0.014)
Foreign-educated immigrant × Degree of horizontal mismatch		-0.026**
		(0.008)
Control Variables	Yes	Yes
Number of Observations	105,183	105,183
Number of Individuals	13,174	13,174

Notes: The control variables (for all models) include gender, race/ethnicity, age, age squared, marital status, years of education, working experience, job tenure, total number of occupational changes, public sector employment, union membership, metropolitan area residency, region, and survey panel. We also controlled for vertical mismatch when modeling horizontal mismatch, and vice versa. Panel B additionally controlled for field of study and occupation.
*** p<0.001, ** p<0.01, * p<0.05.

Section G. A sensitivity analysis by gender

Table G1. Incidence of mismatch by gender

	Model 1	Model 2	
	Vertical Mismatch	Horizontal Mismatch (base category: Horizontal Match)	
		Overmatch	Undermatch
Panel A. Men			
Immigration Status (ref. = native born)			
US-educated immigrant	0.071	-0.061	0.072
	(0.043)	(0.039)	(0.042)
Foreign-educated immigrant	0.126***	-0.064**	0.085***
	(0.026)	(0.024)	(0.025)
Number of Observations	55,344	55,344	
Number of Individuals	6918	6918	
Panel B. Women			
Immigration Status (ref. = native born)			
US-educated immigrant	-0.025	0.024	0.055
	(0.047)	(0.045)	(0.053)
Foreign-educated immigrant	0.110***	-0.009	0.102**
	(0.032)	(0.029)	(0.031)
Number of Observations	51,176	51,176	
Number of Individuals	6397	6397	

The control variables (for all models) include race/ethnicity, age, age squared, marital status, years of education, working experience, job tenure, total number of occupational changes, public sector employment, union membership, metropolitan area residency, region, and survey panel. Additionally, in Model 1 that estimates vertical mismatch, we also controlled for horizontal mismatch; in Model 2 that estimates horizontal mismatch, we also controlled for vertical mismatch.
*** p<0.001, ** p<0.01, * p<0.05.

Table G2. Wage penalties of mismatch by gender

	Vertical Mismatch		Horizontal Mismatch	
Hourly Wage (log transformed)	Model 1 Man	Model 2 Woman	Model 3 Man	Model 4 Woman
Immigration status (ref. = native born)				
US-educated immigrant	0.113*	-0.052	0.168**	-0.094
	(0.044)	(0.049)	(0.051)	(0.064)
Foreign-educated immigrant	-0.037	0.007	-0.097**	-0.016
	(0.028)	(0.031)	(0.033)	(0.038)
Vertical Mismatch (VM)	-0.072***	-0.064***		
	(0.012)	(0.010)		
Interaction				
US-educated immigrant × VM	-0.040	0.082		
	(0.064)	(0.075)		
Foreign-educated immigrant × VM	-0.156***	-0.138***		
	(0.035)	(0.037)		
Horizontal Overmatch (HO)			0.034**	0.048***
			(0.011)	(0.012)
Horizontal Undermatch (HU)			-0.095***	-0.085***
			(0.011)	(0.011)
Interaction				
US-educated immigrant × HO			-0.219**	0.138
			(0.069)	(0.087)
Foreign-educated immigrant × HO			0.075†	0.023
			(0.041)	(0.048)
US-educated immigrant × HU			-0.051	0.095
			(0.056)	(0.083)
Foreign-educated immigrant × HU			-0.063†	-0.152***
			(0.033)	(0.040)
Controls	Yes	Yes	Yes	Yes
Number of Observations	55,344	51,176	55,344	51,176
Number of individuals	6,918	6,397	6,918	6,397

Notes: The outcome variable is hourly wage (log transformed). The control variables (for all models) include race/ethnicity, age, age squared, marital status, years of education, field of study, working experience, job tenure, occupation, total number of occupational changes, public sector employment, union membership, metropolitan area residency, region, and survey panel. Additionally, in Model 1 and 2 that estimate vertical mismatch, we also controlled for horizontal mismatch; in Model 3 and 4 that estimate horizontal mismatch, we also controlled for vertical mismatch.
*** p<0.001, ** p<0.01, * p<0.05, † p<0.1.

Section H. A sensitivity analysis of the intersection of nativity and race/ethnicity

In addition to examining the respective roles of race/ethnicity and nativity, we further investigated how these traits combine to shape patterns and consequences of education-occupation mismatch. An intersectional approach suggests that the nexus of race/ethnicity and nativity may create unique configurations of unequal labor market experiences that are multiplicative as opposed to merely additive (De Jong and Madamba 2001; McCall 2001).

To assess these possibilities, we estimated a new set of models including a categorical variable distinguishing among US-born White, White immigrant, US-born Black, Black immigrant, US-born Hispanic, Hispanic immigrant, US-born Asian, and Asian immigrant. We considered these analyses preliminary because of the relatively small sample size of some race-nativity groups (especially given our focus on college graduates). Also due to the constraints of the sample sizes, we could not further disaggregate immigrants of each ethnoracial group by place of degree.

With respect to the negative types of mismatch (vertical mismatch and horizontal undermatch), the results (Table H1 below) demonstrate mostly additive effects of race and nativity, with immigrants faring worse than the US-born within each race/ethnicity and minority workers being more vulnerable to the negative forms of mismatch than their white peers in both the native and foreign-born groups. Specifically, among the US-born, Black and Hispanic Americans were more likely to be in vertical mismatch than White Americans (Model 1); Black Americans were also more likely to be horizontally undermatched (Model 2). Among immigrants, the risk of vertical mismatch was consistently higher for minority immigrants relative to their White immigrant peers; the risk of horizontal undermatch was higher for Hispanic and Asian immigrants than White immigrants. Hispanic immigrants also had a lower likelihood of horizontal overmatch. One exception was the Black immigrant category, which showed an intersectional effect. Unlike other ethnoracial groups, Black immigrants did not suffer from a higher risk of horizontal undermatch than US-born Black counterparts. These results should be interpreted with caution because of the very small sample size of Black immigrants (68 in total, 24 of whom were horizontally undermatched).

As for the wage consequences of education-occupation mismatch (Table H2), the results show some differences in race/ethnicity but lack a clear pattern. Asian immigrants encountered a greater wage loss from vertical mismatch than not only the US-born Whites (the reference category) but also their co-ethnic US-born peers (Model 1). The pattern for horizontal overmatch and undermatch was more complex. In terms of horizontal overmatch, Black native-born Americans had a higher wage premium, while US-born Hispanic garnered a lower wage premium than US-born White. Turning to horizontal undermatch, the disadvantage immigrants face primarily stems from Black and Asian immigrants who suffered greater wage penalties from horizontal undermatch. Asian and Hispanic native-born Americans also suffered greater wage penalties associated with horizontal undermatch.

Overall, these results suggest that both race and nativity-based processes are at play in education-occupation match. We call for future researchers to use a larger sample size to uncover more robust effects of race, nativity, and the interaction between the two.

Table H1. Incidence of mismatch by nativity and race/ethnicity

	Model 1	Model 2	
	Vertical Mismatch	Horizontal Mismatch (base category: Match)	
		Overmatch	Undermatch
Race/ethnicity and Immigrant Status			
(ref. = US-born White)			
White immigrant	0.542***	0.143	0.375**
	(0.132)	(0.126)	(0.125)
US-born Black	0.308**	-0.198	0.389***
	(0.100)	(0.105)	(0.096)
Black immigrant	0.847**	-0.253	0.130
	(0.328)	(0.329)	(0.325)
US-born Hispanic	0.500***	-0.103	0.246
	(0.146)	(0.144)	(0.139)
Hispanic immigrant	1.092***	-0.888**	0.782***
	(0.220)	(0.290)	(0.206)
US-born Asian	0.382	0.256	0.315
	(0.213)	(0.217)	(0.196)
Asian immigrant	0.871***	-0.175	0.279**
	(0.113)	(0.119)	(0.103)
Control Variables	Yes	Yes	Yes
Number of Observations	106,520	106,520	
Number of individuals	13,315	13,315	

Notes: The results are based on random-effects models. The control variables (for all models) include gender, race/ethnicity, age, age squared, marital status, years of education, working experience, job tenure, total number of occupational changes, public sector employment, union membership, metropolitan area residency, region, and survey panel. Additionally, in Model 1 that estimates vertical mismatch, we also controlled for horizontal mismatch; in Model 2 that estimates horizontal mismatch, we also controlled for vertical mismatch.
*** p<0.001, ** p<0.01, * p<0.05.

Table H2. Wage penalties of mismatch by nativity and race/ethnicity

	Model 1 Hourly Wage (log transformed)		Model 2 Hourly Wage (log transformed)	
	b	se	b	se
Race/ethnicity and Immigrant Status (ref. = US-born White)				
White immigrant	-0.015	(0.025)	-0.060	(0.033)
US-born Black	-0.089***	(0.020)	-0.126***	(0.025)
Black immigrant	-0.132*	(0.067)	0.022	(0.081)
US-born Hispanic	-0.074*	(0.030)	0.020	(0.036)
Hispanic immigrant	-0.294***	(0.049)	-0.331***	(0.056)
US-born Asian	-0.017	(0.041)	0.100	(0.053)
Asian immigrant	0.019	(0.022)	0.015	(0.026)
Vertical mismatch (VM)	-0.067***	(0.008)		
Interaction (ref. = Vertical match)				
White immigrant × VM	-0.064	(0.039)		
US-born Black × VM	-0.011	(0.025)		
Black immigrant × VM	-0.016	(0.103)		
US-born Hispanic × VM	0.013	(0.039)		
Hispanic immigrant × VM	-0.076	(0.058)		
US-born Asian × VM	0.121	(0.062)		
Asian immigrant × VM	-0.141***	(0.031)		
Horizontal Mismatch (ref. = Horizontal match)				
Horizontal Overmatch (HO)			0.046***	(0.008)
Horizontal Undermatch (HU)			-0.076***	(0.008)
Interaction (ref. = US-born White × Horizontal match)				
White immigrant × HO			0.051	(0.046)
US-born Black × HO			0.083*	(0.034)
Black immigrant × HO			-0.150	(0.127)
US-born Hispanic × HO			-0.157***	(0.046)
Hispanic immigrant × HO			0.011	(0.088)
US-born Asian × HO			-0.059	(0.072)
Asian immigrant × HO			-0.003	(0.038)
White immigrant × HU			0.030	(0.040)
US-born Black × HU			0.031	(0.029)
Black immigrant × HU			-0.349***	(0.102)
US-born Hispanic × HU			-0.139**	(0.046)
Hispanic immigrant × HU			-0.003	(0.060)
US-born Asian × HU			-0.190**	(0.068)
Asian immigrant × HU			-0.108***	(0.031)
Number of Observations	106,520		106,520	
Number of individuals	13,315		13,315	

Notes: The outcome variable is hourly wage (log transformed). The control variables (for all models) include gender, race/ethnicity, age, age squared, marital status, years of education, field of study, working experience, job tenure, occupation, total number of occupational changes, public sector employment, union membership, metropolitan area residency, region, and

survey panel. Additionally, in Model 2, which estimates the wage penalty of vertical mismatch, we also controlled for horizontal mismatch; in Model 3 that estimates the wage penalty of horizontal mismatch, we also controlled for vertical mismatch.
*** $p<0.001$, ** $p<0.01$, * $p<0.05$.

Section I. A Sensitivity Analysis of the Mechanisms Shaping Nativity Inequality

We differentiated foreign-educated immigrants along several dimensions as additional tests for the potential mechanisms of immigrants' disadvantages. First, we divided foreign-educated immigrants into two categories by the economic status of their origin countries using the United Nations' definition of developed vs. less developed societies (Guterres 2020).[1] This analysis was restricted to the 1996-2004 panel because the SIPP 2008 dataset did not collect detailed information on country of origin (only broad region categories). We obtained results (Panel 1 in Table I) that were similar to Table 5 in that there was a higher incidence of vertical mismatch and horizontal undermatch among immigrants from less developed countries. Additionally, we found that these immigrants were less likely to attain horizontal overmatch than their native born peers. Second, we differentiated between foreign-educated immigrants from English-speaking origin societies and those from non-English-speaking origin countries according to whether English was the official language of their countries of origin (Panel 2 in Table I). We obtained conclusions that were similar to Table 5 in that the relative disadvantages of foreign-educated immigrants largely concentrated on those from non-English-speaking origin countries.

Table I1. Incidence of mismatch by nativity (AMEs are shown)

	Model 1	Model 2	
	Vertical	Horizontal Mismatch (base category: Match)	
	Mismatch	Overmatch	Undermatch
Panel 1. Economic Status of Country of Origin, SIPP 1996-2006			
Immigration Status (ref. = native born)			
Immigrant from developed countries	-0.033	0.031	-0.019
	(0.038)	(0.040)	(0.045)
Immigrant from less developed countries	0.183***	-0.074**	0.112***
	(0.034)	(0.024)	(0.032)
Control variables	Yes	Yes	
Number of Observations	77,168	77,168	
Panel 2. English vs. Non-English Speaking Country of Origin, SIPP 1996-2006			
Immigration Status (ref. = native born)			
Immigrant from English-speaking countries	-0.060	0.034	-0.071
	(0.041)	(0.048)	(0.048)
Immigrant from non-English speaking countries	0.174***	-0.067**	0.128***
	(0.033)	(0.023)	(0.032)
Control Variables	Yes	Yes	
Number of Observations	77,104	77,104	

Notes: The results are based on random-effects models. The control variables (for all models) include gender, race/ethnicity, age, age squared, marital status, years of education, working experience, job tenure, total number of occupational changes, public sector employment, union membership, metropolitan area residency, region, and survey panel. Additionally, in Model 1 that estimates vertical mismatch, we also controlled for horizontal mismatch; in Model 2 that estimates horizontal mismatch, we also controlled for vertical mismatch. The end year reported in the table is the last year of the SIPP panel used in the analysis.
*** p<0.001, ** p<0.01, * p<0.05.

[1] The data are available at: https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/WESP2020_Annex.pdf

References

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies*, *58*(2), 277–297.

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, *68*(1), 29–51.

Blundell, R., & Bond, S. (2000). GMM estimation with persistent panel data: an application to production functions. *Econometric Reviews*, *19*(3), 321–340.

De Jong, G. F., & Madamba, A. B. (2001). A double disadvantage? Minority group, immigrant status, and underemployment in the United States. *Social Science Quarterly*, *82*(1), 117–130.

Duncan, G. J., & Hoffman, S. D. (1981). The incidence and wage effects of overeducation. *Economics of Education Review*, *1*(1), 75–86.

Guterres, A. (2020). World Economic Situation Prospects 2020. New York, NY: United Nations.

Kerckhoff, A. C. (1996). Building conceptual and empirical bridges between studies of educational and labor force careers. In A. C. Kerckhoff (Ed.), *Generating Social Stratification: Toward a New Research Agenda* (pp. 37–56). Boulder, CO.: Westview Press.

Kiker, B. F., Santos, M. C., & de Oliveira, M. M. (1997). Overeducation and undereducation: Evidence for Portugal. *Economics of Education Review*, *16*(2), 111–125.

Lu, Y., & Li, X. (2021). Vertical Education-Occupation Mismatch and Wage Inequality by Race/Ethnicity and Nativity among Highly Educated US Workers. *Social Forces*, *100*(2), 706–737.

McCall, L. (2001). Sources of Racial Wage Inequality in Metropolitan Labor Markets: Racial, Ethnic, and Gender Differences. *American Sociological Review*, *66*(4), 520–541.

McGuinness, S. (2006). Overeducation in the labour market. *Journal of Economic Surveys*, *20*(3), 387–418.

Ortiz, L., & Kucel, A. (2008). Do fields of study matter for over-education?: The cases of Spain and Germany. *International Journal of Comparative Sociology*, *49*(4–5), 305–327.

Roodman, D. (2009). How to do xtabond2 : An introduction to difference and system GMM in Stata. *The Stata Journal*, *9*(1), 86–136.

Rumberger, R. W. (1981). *Overeducation in the US labor market*. New York: Praeger Publishers.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.