ONLINE APPENDIX A: Methodological Appendix

As I mention in the main paper, for the review of articles, I gave my research assistants instructions to create a list of articles that met four criteria: 1) It was an article published from 2008 to 2018 in Ethnic and Racial Studies (ERS), The International Migration Review (IMR), or the Journal of Ethnic and Migration Studies (JEMS); 2) it was a study of the integration of immigrants, broadly construed and that contained the words "integration", "incorporation" or "assimilation; 3) it was a study conducted at the subnational level and 4) it was not a theoretical piece.

In this paper, subnational level is defined as any spatial unit, administrative, statistical, or vernacular, that is smaller than that nation-state. This includes administrative units like municipalities; statistical units like census tracts and European NUTS and common-use definitions like Accra and New York. After a final list was compiled, the research assistants and I coded articles in the 2018 editions of ERS and JEMS independent of each other. When this was finished, we reconvened to check for mistakes, validate for consistency, and discuss emergent categories. Although the original coding had only foreseen coding for sites of study and ethnic groups, it quickly became apparent that we should also keep track of other information, including the scale the study was being conducted at (Scale), the number of sites (N_Sites), a dummy for whether the text included the words or variants of the words integration, incorporation or assimilation as they relate to immigrants in the body of the text (excluding references) (Incorporation), the Nation-State the study was conducted in or whether it was a transnational study (Nation) and a dummy for whether the study was a policy study. Ultimately, many of these variables were not used in the final analysis, but the Scale, N_Sites, Incorporation and Nation, proved eminently useful categories for the analysis.

Ultimately, scale was coded into four sub-national categories in this analysis: Region, Settlement, Neighborhood and Organizations. Defining the boundaries of these categories is not clear cut, but the idea is to create categories that are nested within each other and of a geographically different size. Region refers to large areas that encompass many human settlements and surrounding uninhabited space. An example of this geographic entity is the American West or NUTS1 designation in Europe. Settlements are defined as spatially-bounded local places with human populations. This is a definition that Brenner and Schmid (Brenner and Schmid 2015) explicitly argue against, but it is preferable to rural/suburban/urban categories which have no clear distinction and obfuscate more often than they clarify (Brenner and Schmid 2014; Gans 2009). Since I am interested in variation in local contexts of integration, I wanted a term that would designate a local scale but that was universal enough to capture the varied types of places immigrants live. Settlement, following Gans suggestions, was a good fit and was preferred over local because settlements are conceptually less flexible in terms of scale. Neighborhoods refer to geographic areas within settlements, and this is the definition used here. This includes things like ethnic enclaves or Little Italy. Finally, organizations follow Small and McDermott's (2006) definition as an "establishment that has a physical location and offers services or sells goods basic to day-to-day living." These are the most granular scale that is coded because more granular scales would be at the level of individuals. As was noted in the main article, texts were coded with an eye towards remaining faithful to the intention of the authors. This was usually quite easy, but sometimes required some inference. This was a recurrent issue for many of the variables, but was especially the case when coding the scale category since references to scales could vary within articles.

The compilation of a list of books was conducted differently. I used "Publish or Perish" (PoP), a citation scraper and aggregator (Harzing 2010), to extract a list of highly cited texts that were published from 2008 to 2018. Using the search term "immigration OR immigrant OR immigrants OR migrant OR migration OR migrants" in the title field, texts were scraped 1,000 books at a time limited by year.[1] This returned the thousand most-cited texts containing the search terms in the title for a given year. Searching in the title field was more effective than a general search because this was more likely to return books than the latter which often returned journal articles.

This operation was repeated ten times for every year of the analysis until a data frame containing 10,000 observations was created. To make the coding manageable, I removed all observations that did not have at least 300 citations or observations were journal articles, natural science texts, reference texts or readers, works of fiction or were written in a language other than English. This substantially reduced database was then coded using the same conventions used to code the journal articles.[2] Since many of the highly cited books were edited volumes with stand-alone book chapters, each chapter was coded individually in these cases. In all, after filtering out books of immigrant integration at the national level or migration theory, 78 books and book chapters were coded.

After the construction of the book and article database was complete, it was imported into R and underwent second-cycle coding (Miles, Huberman, and Saldaña 2014) to create consistent and analogous scale categories (Scale1) and site categories (Site1). Additionally, a Site2 category was created to aggregate observations at the neighborhood and organizational level into the settlement

---

[1] This is the maximum amount of texts that can be extracted per cycle. The 1,000 shown are the most-cited results for that search (Harzing 2016).

[2] The incorporation variable was harder to code for books where the texts were not digitized and searchable. Only a small number of books had this issue. For these books, a best attempt was made to find these terms.

those observations were conducted in (e.g. Organizations in Los Angeles or East Los Angeles became Los Angeles). It was from these second-cycle codes that the tables found in the article were compiled.

After the data was coded, 2010 population and capital city information were added. Tables 1A and 4A offer additional summary of this data. Population data for settlements was taken from the 2018 Revision of the World Urbanization Prospects produced by the United Nations Department of Economic and Social Affairs, Population Division. Since this data only covers settlements with more that 300,000 people, population estimates for remaining sites were taken directly from the closest 2010 census estimates available in each country. When possible, population ranking was derived using the World Urbanization prospects. When the list of top three cities in a country was not available (for example when the third largest city has less than 300,000 people) this ranking was taken from Wikipedia. Capital city information was also taken from Wikipedia. Provincial capitals are defined as the administrative seat of the first-order administrative division in a nation. This includes divisions like American States, Polish voivodeships and Turkish Provinces.

The final database used for the analysis contained 1,070 text-observations. In total, 436 articles and 78 books and book chapters were reviewed. The original data of the articles and books that were coded for this study, as well as the scripts for the analytical coding, are available upon request. There are possibly a hundred other ways this database could have been constructed, and I make no claims that this data is "representative" of migration texts. Rather, I am intent on showing that geographic clustering is an existing trend and that it is present in enough notable texts to warrant reflection. For this purpose, a sub-sample of prominent texts, even if unrepresentative will suffice. Especially if, as I have documented, these are present in three major journals and highly cited books. Future versions of this analysis should consider expanding the scope to include articles in

other languages such as Spanish, Mandarin, French, and German, and including the set of journals

and books reviewed to those published in the other outlets, especially in the global South.

REFERENCES

Adler, Nancy J., and Anne-Wil Harzing. 2009. "When Knowledge Wins: Transcending the Sense and Nonsense of Academic Rankings." *Academy of Management Learning & Education* 8(1):72–95.

Brenner, Neil, and Christian Schmid. 2014. "The 'Urban Age' in Question." *International Journal of Urban and Regional Research* 38(3):731–55.

Brenner, Neil, and Christian Schmid. 2015. "Towards a New Epistemology of the Urban?" *City* 19(2–3):151–82.

Gans, Herbert. 1997. "Best-Sellers by Sociologists: An Exploratory Study." *Contemporary Sociology* 26(2):131.

Gans, Herbert. 2009. "Some Problems of and Futures for Urban Sociology: Toward a Sociology of Settlements." *City & Community* 8(3):211–19.

Harzing, Anne-Wil. 2010. *The Publish or Perish Book : Your Guide to Effective and Responsible Citation Analysis*. Melbourne: Tarma Software Research Pty Ltd.

Harzing, Anne-Wil, and R. van der Wal. 2008. "Google Scholar as a New Source for Citation Analysis." *Ethics in Science and Environmental Politics* 8:61–73.

Jacobs, Jerry. 2016. "Journal Rankings in Sociology: Using the H Index with Google Scholar." *The American Sociologist* 47(2):192–224.

Longhofer, Wesley, Shannon Golden, and Arturo Baiocchi. 2010. "A Fresh Look at Sociology Bestsellers." *Contexts* 9(2):18–25.

Miles, Matthew B., A. M. Huberman, and Johnny Saldaña. 2014. *Qualitative Data Analysis: A Methods Sourcebook*.

van Raan, Ton. 2010. "Bibliometrics: Measure for Measure." *Nature* 468:763.

Sirkeci, Ibrahim, Jeffrey Cohen, and Andrej Přívara. 2017. "Towards a Migration Letters Index: The Most Influential Works and Authors in Migration Studies." *Migration Letters* 14(3):397–424.

Small, Mario Luis, and Monica McDermott. 2006. "The Presence of Organizational Resources in Poor Urban Neighborhoods: An Analysis of Average and Contextual Effects." *Social Forces* 84(3):1697–1724.

TABLES

Table 1A. Settlements Studied by Population (Percentage)

Population	N	%
500,000 or less	229	27.29
500,000 to 1 million	80	9.54
1 million to 1.5 million	87	10.37
1.5 million to 2 million	50	5.96
2 million to 2.5 million	32	3.81
2.5 million to 3 million	24	2.86
3 million to 3.5 million	43	5.13
3.5 million to 4 million	21	2.50
4 million to 4.5 million	22	2.62
4.5 million to 5 million	29	3.46
5 million or more	222	26.46
Total	839	100.00

Table 2A. Texts that Study At Least One Large City

	1st Largest		2nd Largest		3rd Largest		Top 3	
	N	%	N	%	N	%	N	%
City	230	43.9	113	21.6	64	12.2	311	59.4
Other	294	56.1	411	78.4	460	87.8	213	40.6
Total	524	100.0	524	100.0	524	100.0	524	100.0

Table 3A. Texts that Study At Least One Capital City

National			Provincial			Either National or Provincial		
	N	%		N	%		N	%
Capital	192	36.6	Capital	268	51.1	Capital	288	55
Other	332	63.4	Other	256	48.9	Other	236	45
Total	524	100.0	Total	524	100.0	Total	524	100

Table 4A. Settlement Study Frequency (More Than Once)

	Site Population	N
London	8044433	46
Los Angeles	12160151	35
Amsterdam	1064617	27
New York City	18365262	25
Madrid	5787392	20
Rotterdam	1002773	17
Barcelona	4933548	16
Berlin	3450395	16
Stockholm	1359972	16
Toronto	5499233	16
Chicago	8615834	11
Paris	10460118	10
The Hague	636704	10
Vienna	1730649	10
Boston	4184792	9
Oslo	897916	9
San Francisco	3282535	9
Vancouver	2278216	9
Birmingham	2428921	8
Lisbon	2811563	8
Melbourne	3932257	8
Miami	5517777	8
Milan	3056107	8
San Diego	2964154	8
Hong Kong	7025221	7
Rome	3957247	7
Sydney	4165739	7
Antwerp	992577	6
Malmö	278087	6
Utrecht, Netherlands	466013	6
Washington, DC	4604365	6
Brussels	1926348	5
Copenhagen	1192357	5
Houston	4976162	5
Singapore	5074252	5

Table 4A. Settlement Study Frequency (More Than Once)

	Site Population	N
Basel	517610	4
Dublin, Ireland	1100238	4
Glasgow	1616798	4
Leeds	474632	4
Orange County, California	3010232	4
San Antonio	1770572	4
Shanghai	20314309	4
Turin	1736536	4
Aarhus	306650	3
Athens	3170434	3
Dordrecht	97690	3
Eindhoven	328683	3
Fort Lauderdale	165521	3
Frankfurt	681370	3
Gothenburg	545763	3
Hamburg	1775365	3
Johannesburg	4249354	3
Kampala, Uganda	2015559	3
Krakow	759044	3
Leicester	504454	3
Linz	189889	3
Moscow	11461264	3
Oxford	151906	3
Perth	1722319	3
Philadelphia	5449069	3
Phoenix	3649247	3
Riverside County, California	2189641	3
Strasbourg	449931	3
Tokyo	36859626	3
Warsaw	1702689	3
Zurich	1257280	3
Aalborg	197426	2
Accra	2060076	2
Almere	190655	2
Alphen aan den Rijn	72680	2

Table 4A. Settlement Study Frequency (More Than Once)

	Site Population	N
Atlanta, GA	4544198	2
Beijing	16441252	2
Bergen op Zoom	66074	2
Bradford, United Kingdom	349561	2
Buenos Aires	14245871	2
Delft	97690	2
Enschede	157838	2
Finnmark County	73936	2
Florence	692620	2
Garden City, Kansas	26658	2
Guangzhou	10278197	2
Hargeisa, Somalia	606912	2
Hoogezand-Sappemeer	34814	2
Istanbul	12584885	2
Jerusalem	781421	2
Lima	8919759	2
London, Ontario	471805	2
Lund, Sweden	111792	2
Lyon	1576653	2
Manchester	2537770	2
Mexico City	20136681	2
Montgomery County, Maryland	971777	2
Montreal	3896932	2
Nairobi	3236589	2
New Delhi	21987895	2
Oldham, United Kingdom	96555	2
Ottawa	1218013	2
Prato	184885	2
Raleigh-Durham	895794	2
San Bernardino	1944699	2
San Jose	1667771	2
Santa Maria, California	99553	2
Southampton	851089	2
St.Paul	2657786	2
Tallinn	394129	2

Table 4A. Settlement Study Frequency (More Than Once)

	Site Population	N
Thessaloniki	806139	2
Tiel	41181	2
Uppsala, Sweden	199898	2
Ventura County, California	823318	2
Wroclaw	632652	2