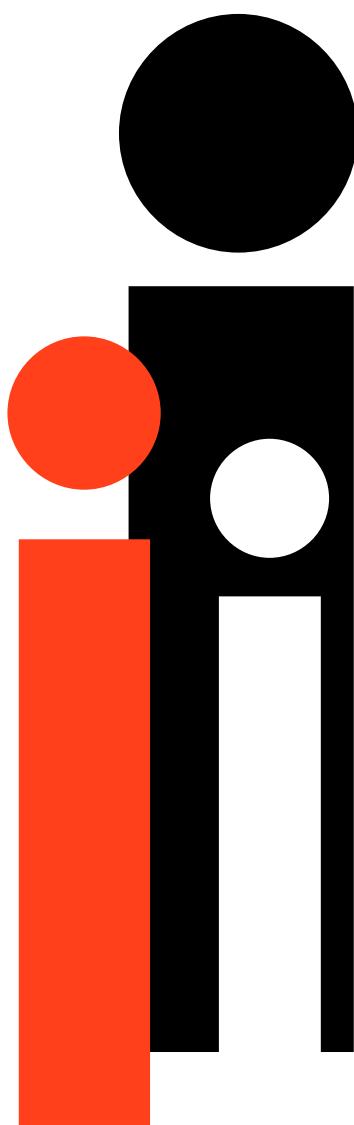


Who assimilates? Statistical artefacts and intergenerational mobility in immigrant families

Renee Reichl Luthra
Institute for Social and Economic Research
University of Essex

Thomas Soehl
Dept of Sociology
McGill University

No. 2014-28
August 2014



**INSTITUTE FOR SOCIAL
& ECONOMIC RESEARCH**

Non-Technical Summary

One in five US residents under the age of 18 has at least one foreign-born parent. Although the distribution of immigrants in terms of human capital is bimodal – the foreign born have disproportional concentrations at the highest and lowest skill levels – it is especially the large group of immigrants with little formal education that raises concerns about the impact of immigration on social inequality. Whether their educational disadvantage will persist and shape stratification in the over the long run is determined by the degree of intergenerational educational transmission: to what extent do foreign born parents pass on their educational advantage or disadvantage to their children?

This project utilizes new data on the children of immigrants from 18 different origin countries in four US metropolises, assessing highly influential estimates of immigrant intergenerational mobility that are based on aggregate data sources. We show that aggregation bias strongly inflates estimates of the relationship between immigrants' educational attainment and the educational attainment of their children. Compared to natives, the educational transmission process between parent and child is much weaker in immigrant families. A number of group-level processes, such as societal discrimination, ethnic segregation, or ethnic networks, may render group characteristics more important predictors of second generation educational attainment than parental education. We emphasize the importance of a clear analytical and empirical distinction between group- and individual level processes in research on immigrant assimilation.

**Who assimilates?
Statistical artefacts and intergenerational mobility in immigrant families**

Renee Luthra
ISER, University of Essex

Thomas Soehl
Dept of Sociology, McGill University

Abstract:

This paper assesses estimates of immigrant intergenerational mobility that are based on aggregate data sources. We show that aggregation bias strongly inflates estimates of the relationship between immigrants' educational attainment and the educational attainment of their children. Compared to natives, the educational transmission process between parent and child is much weaker in immigrant families. A number of group-level processes, such as societal discrimination, ethnic segregation, or ethnic networks, may render group characteristics more important predictors of second generation educational attainment than parental education.

Keywords: immigration, intergenerational mobility, education

JEL: I210 ; J150 ; J620.

Corresponding author:

Renee Luthra, ISER, University of Essex
Wivenhoe Park, Colchester, CO23SQ, UK
rnluthra@essex.ac.uk

Acknowledgments:

This work was supported by the Economic and Social Research Council (ESRC) through the Research Centre on Micro-Social Change (MiSoC) (award no. RES-518-28-001). The authors also gratefully acknowledge support from the Russell Sage Foundation and the Spencer Foundation Small Grants program.

Introduction

The initial members of the “new” immigration wave following the 1965 Immigration and Nationality Act, originating from Latin America and Asia, have now settled and their US born children have come of age. Although the distribution of immigrants in terms of human capital is bimodal – the foreign born have disproportional concentrations at the highest and lowest skill levels – it is especially the large group of immigrants with little formal education that raises concerns about the impact of immigration on social inequality. As the children of immigrants currently comprise more than 20% of the US population under the age of 18, the question to which extent this population will inherit the educational characteristics of their parents has significant consequences for the immediate and long-term future of ethnic stratification in the United States.

Until recently, the answer to this question has been difficult to obtain. Although intergenerational mobility has occupied a central position in quantitative sociological inquiry for several decades (Blau and Duncan 1967; Hout and DiPrete 2006; Mare 1981), representative, large-scale data identifying the educational attainments of immigrants and their adult children have been scarce. Lacking individual level characteristics of immigrant parent and adult child, researchers have relied instead on aggregate data sources, linking national-origin estimates of the educational attainment of immigrants to national origin or self reported ethnicity groupings observed among the children of the foreign born in later survey years (Borjas 1993; Borjas 2006; Card 2005; Card, DiNardo, and Estes 2000; Park and Myers 2010; Smith 2003). Estimates of the intergenerational transmission of educational attainment in immigrant families using these methods consistently fall between 0.3 and 0.4, and have been interpreted to indicate that intergenerational mobility is similar for immigrants and natives, and that intergenerational mobility has been fairly consistent across immigrant cohorts (Card, DiNardo, and Estes 2000)¹. This conclusion is highly influential in the economics and sociology of migration literatures. At this moment the papers by Card and colleagues alone have been cited over 600 times and the regression coefficient estimate of 0.4

¹ The original version of this paper was published as: Card, David, John DiNardo, and Eugena Estes. 1998. "The More Things Change: Immigrants and the Children of Immigrants in the 1940s, the 1970s, and the 1990s." *NBER Working Paper 6519*.

currently serves as a benchmark for comparative estimates from alternative datasets and other countries.

In this paper, we evaluate these influential estimates using recent individual level data on the educational attainment of the children of immigrants and their parents. We find substantial discrepancies between estimates of intergenerational mobility using aggregate and individual level data: using family level parent-child dyads, the regression coefficient of children's years on parents' highest years of education is 0.13 on average for immigrants. It is 0.2 or lower in about three quarters of the national origin groups in our surveys and in many cases below 0.1. In contrast, when aggregating our data and using weighted averages of national origin groups, as has been done in prior research, we find much higher estimates - an association between foreign-born parents' and their children's education of 0.43.

These results are robust across different metropolitan-level and national-level data sources, as well as to the specifications used in identifying the second generation or whether mother's, father's or the highest level of parental education is used.-We further test the sensitivity of our results to reporting error at the individual level, triangulating multiple reports of parental education in a latent variable model. The results suggest that although reporting error attenuates individual level estimates of intergenerational mobility, the resulting bias is relatively slight. We argue that due to aggregation bias estimates of the intergenerational association of education among immigrant *groups* should not be interpreted as estimates of intergenerational transmission and certainly not be used as a benchmark for individual level studies. We conclude with the implications of these findings for previous evaluations of the assimilation trajectories of immigrants.

Assimilation and Intergenerational Mobility

Sociologists of migration have long been interested in intergenerational change among immigrants, writing extensively on the earlier “great wave” of migration at the turn of the century (Gordon 1964; Park 1930; Warner and Srole 1945). The influx of Catholics and Jews from Eastern and Southern Europe, alongside already existing Asian minorities and

African origin involuntary migrants, resulted in a society complexly stratified along racial, religious, and national-origin lines. Observing this stratification, the original formulations of assimilation theory conceived of assimilation as a group-level process, predicting a sequence of improving *group* relations with the disappearing of ethnic groups as its endpoint. Even Gordon's influential treatise on the subject, while introducing a multi-dimensional approach to assimilation, is framed as a corrective to the lack of "research and theoretical attention to the nature and implications of American communal group life" (1964:5).

These approaches were extraordinarily productive, guiding immigration research for the better part of a century, yet they do not clearly delineate between individual and group level processes. Assimilation is seen as a convergence of immigrant *groups* towards the "core", and the disappearance of prejudice and discrimination. At the same time it encompasses processes that are clearly individual in nature such as intermarriage, shifts in participation and identification.

It is the achievement of Alba and Nee's reformulation of assimilation theory (1997, 2003) to clearly establish an analytical model in which individual striving for socio-economics advancement is the central mechanism behind assimilation, with individuals (and their families) as the actors and key analytical units. On the individual level socio-economic mobility, intermarriage, and residential assimilation of a (multi-generational) migrant population will determine to what extent national-origin characteristics and ethnic identifications persist into later generations. Individuals may "leave" the ethnic group (boundary crossing) by changing their mode of identification, moving away from ethnic enclaves, and leaving ethnic occupations niches. On a collective level, the salience of ethnicity may decline across time (boundary blurring or boundary shifting) or vary across different domains of life. Ethnic group dynamics certainly matter in this framework – in fact the absence of strong institutionalized forms of ethnic closure are prerequisites for individual level processes of assimilation to work - but they are variable and individuals rather than ethnic groups are the constituent elements of society.

In parallel with this sociological literature, economists have developed an empirical literature on immigrant intergenerational mobility with a focus on the convergence (or lack thereof) of immigrant populations towards the mainstream in terms of socio-economic characteristics across generations. But just as early versions of assimilation theory, this literature at times wavers between a group-based analysis of assimilation and individual level interpretations.

Some analyses clearly separate individual level from group level influences. Borjas (1992, 1995) for example estimates both the family level transmission and the influence of “ethnic capital” – measured as group averages of measures such as education or occupational status. Yet, when drawing conclusions the estimates of family level- and group level processes are lumped together to argue that there are substantial linkages across generations (e.g. Borjas 1992: 139). It is not clear, however, how broadly these group level “effects” apply and to what extent they hold over generations. After all exit from the group as a result of processes such as socio-economic mobility, residential assimilation or intermarriage and resulting shifts in identification, is a key outcome of assimilation.² Yet, most analysis in that vein, including those of Borjas, define group membership via self-identified ethnicity or ancestry. Thus the individuals who are most assimilated and thus “lost to the group” do not enter the estimation. The small samples of second-generation and ethnic minorities that can be identified in suitable data sources such as the General Social Survey (GSS) or National Longitudinal Study of Youth (NLSY) are another limitation for this research.

To obtain a broader empirical base other research has relied on the US Census. As these data do not include the identification of actual parent and child dyads, this line of work ignores the individual level altogether and takes ethnic group averages as the source of data. One popular approach is to regress the average years of education of second generation national origin groups on the average years of education of immigrants from

² A substantial literature has shown ethnic identification is malleable and responsive to local as well as global shifts in social context. For examples see: Nagel, Joane. 1995. "American Indian Ethnic Renewal: Politics and the Resurgence of Identity." *American Sociological Review* 60:947–965. Waters, Mary C. 1999. *Black identities: West Indian immigrant dreams and American realities*. New York: Russell Sage Foundation.

the same national origins from a Census 20 to 30 years earlier. In this case, the analytic approach corresponds to an “old school” conception of assimilation that takes ethnic groups as the unit of analysis.

Using this method a series of highly influential articles by Borjas (1993; 1994) demonstrated that links in educational attainment and reported wages between first and second generation immigrants of the same ethnic origins, and even between first and third generation immigrants, are strong and significant, suggesting intergenerational *immobility* and a slow process of assimilation towards the US population mean. In subsequent work using similar methods Card and colleagues (2000, 2005) find a coefficients of about 0.4 and 0.3 respectively, which is similar to a coefficient estimated from parent-child dyads in the native population using the General Social Survey (GSS) (Card 2005, footnote 30). This finding led to the conclusion that “... the intergenerational transmission of education is about the same for families of immigrants as for other families in the US” (p.319). Similarly, this work has found that the degree of intergenerational transmission is similar to that of earlier immigrant cohorts (Card 2000; Borjas 2006).

Despite their lack of individual level information, this series of articles and the approach they apply are exceptionally influential, cited in virtually every subsequent article on immigrant intergenerational mobility, and taking a central position in recent reviews of the mobility literature. At this moment the papers by Card and colleagues alone have been cited over 600 times and the regression coefficient estimates of 0.3 and 0.4 currently serve as benchmarks for comparative estimates from alternative datasets and other countries.

Across a number of countries, studies using aggregate data (Dustmann and Glitz 2011; Smith 2003) find consistently much higher estimates of transmission than those using comparable micro/family level data. At the same time, studies that have estimates of individual level, parent to child transmission, we see that these are consistently lower in immigrant families than for those with native born parents (Aydemir, Chen, and Corak 2008; Bauer and Riphahn 2006; Borjas 1992; Riphahn 2003) in other cases the transmission estimates for immigrant families are statistically not significant while there is significant association in levels of education across generations

among natives (Dustmann 2008; Gang and Zimmermann 2000; Nielsen, Roshholm, Smith, and Husted 2003).³ Two recent papers, an OECD review of intergenerational mobility studies by d'Addio (2007:Box 10) and Dustmann and Glitz (2011) have already noted the discrepancies in different estimates but do not address the source of the confusion.

As we will show in the remainder of this paper, the conceptual oscillation between assimilation as a narrowing of group differences and individual level processes has a methodological cousin: ecological fallacy, or biased – inference about micro-level processes, such as intergenerational transmission within individual families, from aggregate level data.. As we summarize in the next section, these cross-level inferences are valid only under a very specific set of conditions. While some of the early Census research used careful formulations to not attribute individual level processes to aggregate level findings (Borjas 1993, 1994) or at least discussed these limitations (Card et al. 2000: 251), they have since been taken wholesale as estimates of the intergenerational transmission process and are used as a point of comparison for individual level studies of immigrant intergenerational mobility.

Aggregate data and individual level processes: Aggregation bias and ecological fallacy

Robinson's (1950) path breaking article on ecological correlations drove home the point that aggregate data, in most cases, can not be used to draw inferences about individual level phenomena. A key example from his article is the use of aggregate data to determine the relationship between literacy rates and immigration. Although immigrants at the turn of the century had higher rates of illiteracy than the native born population, when looking at a correlation between illiteracy rates and immigrant share by state the correlation is negative (-0.53). The reason was that immigrants settled overwhelmingly in the industrialized states where literacy rates were higher than in the rural southern states.⁴ This article, cited over 3000 times, sparked a veritable cottage industry of methodological

³ A table summarizing these studies in more detail is available from the authors upon request.

⁴ The other example in the article shows that aggregate data dramatically overestimates the illiteracy rates among African-Americans.

research that examined the problem and established the conditions under which valid individual level inferences can be drawn from aggregate data.

It is now established that when looking at correlations, as Robinson did, coefficients will necessarily be higher in magnitude when making inferences based on aggregate data if observations are grouped on an external grouping variable – even absent confounding factors. However, regression analysis may, under the right conditions, still provide accurate cross level inference (Firebaugh 1978; Goodman 1953; Hammond 1973). What these conditions are has been conceptualized in a variety of ways. One way of stating the requirement is that the relationship between variables on the individual level is the same across units of aggregation (Goodman 1953, Hammond p.765). In our case this would mean that the relationship between foreign born parental education and second-generation education does not vary across immigrant origin groups. Groups with low levels of parental education must represent all individual immigrants with low levels of education. When only aggregate data is available, this of course cannot be evaluated empirically but has to be assumed.

Another way to frame the requirement is in terms of omitted variables bias. A standard assumption in any regression is that the error term is uncorrelated with the independent variables – this correlation representing an omitted variable that affects the outcome of interest. When using aggregate data this means that the mean of the error terms is uncorrelated with the means of the independent variables. It is easily possible that in the same data this requirement is satisfied at the individual level, but not when using aggregate data (Jargowsky 2004:9). For example if the external grouping variable is associated with the outcome variable aggregation itself will introduce an omitted variable. And if the grouping variable is related to a variable not included in the (individual level) model it can exacerbate omitted variable bias. To take the example from Robinson above – the grouping variable here are states - since those states with lower illiteracy rates (outcome variable) had higher shares of immigrants, in a regression of state illiteracy rate on immigrant share, the coefficient is negative. Thus a naive interpretation could be that immigrants have lower illiteracy rates than the native population – an obviously false conclusion in this context.

One way to formally state the requirement is that, for an estimate from aggregate data to be equivalent to those from individual level data, in a (hypothetical) individual level model, the mean of the independent variable can provide no additional information on the outcome variable. We illustrate this with our simple bivariate case – regressing the educational achievement y (measured in years) of individuals i in group j on an intercept α and the educational standing (in years) of their parents x . Using the notation below, the coefficient indicating the effect of the mean education level (and all associated, “omitted variables”), β_2 must be equal to zero for this assumption to hold (see also Firebaugh 1978, 560).

$$y_{ij} = \alpha + \beta_1 x_{ij} + \beta_2 \bar{x}_j + \varepsilon_{ij} \quad (1)$$

It is easy to see how this condition may be violated in models of intergenerational mobility in immigrant families. There are several ways in which we can imagine relationships between the grouping variable – immigrant national origin – and both parental educational attainment and with respondent’s educational attainment. Or put differently how the mean level of parental education in a group is associated with second generation outcomes above and beyond parental education.

However, when using aggregate data we can not differentiate the two. Any regression of group level means implicitly measures *gross*, rather than net transmission rates – in other words, the effect of the individual’s parents characteristics as well as the average characteristics for the group as whole (Borjas 1995:374; Jargowsky 2004). In an aggregate level regression, the individual level regression (1) above becomes:

$$\begin{aligned} \bar{y}_j &= \alpha + \beta_1 \bar{x}_j + \beta_2 \bar{x}_j + \varepsilon_j \\ &= \alpha + (\beta_1 + \beta_2) \bar{x}_j + \varepsilon_j \end{aligned} \quad (2)$$

And thus we can no longer separately identify β_1 and β_2 . When using aggregate data, the coefficient of the average outcome of the second generation regressed on the average outcome of the immigrant parent of the same origin contains *both the intergenerational transmission coefficient and the group level effects*.

Figure 1 illustrates how aggregation can cause bias using data from two groups, Mexican and Chinese origin respondents, from a recent second-generation survey in Los Angeles. A detailed description of data and analysis is in the following section. The regression lines based on individual data for both groups (grey for Mexicans, black for Chinese) show a relatively weak relationship between parental and second-generation education as measured in years. The coefficients are 0.13 and 0.14 respectively. Once aggregating and using group mean to fit a regression we get a much steeper line with a large slope coefficient of almost 0.6.

It is beyond the scope of this paper to give an exhaustive account of the possible social processes often referred to as “group effects” that may account for aggregation bias in estimates of immigrant intergenerational mobility. Sorting into occupations and neighborhoods are certainly one important part of the story. Especially immigrants from less developed countries are concentrated in lower paying occupations and often live in segregated neighborhoods (Cutler, Glaeser, and Vigdor 2008; Piore 1980), which in turn impacts the educational opportunities for their children. Even in the absence of receiving country discrimination, “ethnic social capital,” the social ties that are a central part of the migration and settlement process (e.g. Massey 1998; Waldinger and Lichter 2003) shape the assimilation trajectories of the next generation and beyond (Tilly 1998). Group effects may also be mediated through neighborhood institutions. Given clear patterns of residential concentration along national origin lines, the quality of schools or other neighborhood institutions is a probable mechanism by which average economic and human capital resources of an ethnic group affect the educational outcomes of the second generation (Borjas 1992, 1995). How these same mechanisms can also help migrant families achieve disproportional mobility given their background has been shown in the example of Sikhs in California (Gibson 1988), religious networks of Vietnamese in New Orleans (Bankston and Zhou 1995) or the positive effect of cross-class ethnic solidarity among the Chinese in New York City (Kasinitz, Mollenkopf, Waters, and Holdaway 2008).

On the other hand, absent these resources, ethnic social networks can compound individual disadvantage. Poor quality neighborhoods and exclusion from information

channels about how to navigate receiving country institutions can limit access to educational and economic opportunity. More generally, segmented assimilation theory argues for the importance of context of reception and ethnic social capital as central factors for the prospects of today's second generation (Portes and Rumbaut 2001).

Finally, we also expect a weaker individual level relationship between immigrant parents and their children because the educational attainment of immigrant parents, who are largely educated outside the United States, may not be a good indicator for predicting a family's educational success in the United States. Especially in countries where education is expensive or opportunities are not allocated according to ability or ambition (or less so than in the US), years of education may be a poorer measure of parental human capital, educational values, and intelligence for immigrants as compared to someone educated in the US, and therefore a weaker indicator for the actual mechanisms of intergenerational transmission. A related issue is selective migration, which may also weaken the observed relationship between educational attainment and ability for immigrant parents. If only the most (un)motivated and (un)able migrate independent of their educational characteristics, the observed relationship between parental and child's education will be attenuated relative to the "true" relationship absent immigrant selection.

Intergenerational transmission of education in migrant families: comparing micro level and macro-level data approaches.

We now turn to several recently released datasets that provide information on the educational attainment of second-generation adults and their parents from a variety of national origin groups, enabling us to estimate intergenerational transmission of education using both individual level and aggregate level data. In total we utilize data from four different surveys collected over the last decade. Three of these surveys sampled second-generation respondents in four different metropolitan areas in the United States: The Immigration and Intergenerational Mobility in Metropolitan Los Angeles (IIMMLA) survey, the Immigrant Second Generation in Metropolitan New York survey (ISGMNY) and the Children of Immigrants Longitudinal Survey (CILS) which surveyed the children

of immigrants in San Diego and Miami. In addition we rely on a nationally representative survey of young adults that provides substantial samples of several national origin groups - the National Education Longitudinal Study (NELS). In contrast to previous research (Borjas 1992, 1995) we do not draw on the National Longitudinal Study of Youth (NLSY), because this study lacks sufficient numbers to examine second generation youth at the national origin level.

Data

Second generation surveys – IIMMLA, ISGMNY, CILS: These three surveys were collected in the last decade to ascertain the assimilation trajectory of the children of post 1965 immigrants. While differing a bit in the exact battery of questions asked and the parameters of their sampling frames all three provide extensive information on respondents' educational trajectory as well as the education background of their parents. Though these surveys have the disadvantage of not being nationally representative, they employ quota sampling of a variety national origin groups and thus provide significant sample sizes for national origin groups that an ordinary nationally representative sample could not capture.

- ISGMNY, conducted in 1998 and 1999, entailed a telephone survey, interviewing 3,415 young adults, aged 18 to 32 in New York City and its surrounding suburbs. The survey targeted second generation Chinese, Dominicans, Russian Jews, West Indians and Central Americans from Colombia, Ecuador and Peru. It also includes comparison groups of native Blacks, Puerto Ricans and non-Hispanic Whites.
- Also a telephone survey, IIMMLA was conducted in 2004 and collected approximately 4500 interviews with young adults aged 20 to 39 in the Los Angeles Metropolitan area – comprising Los Angeles, Orange, Ventura, Riverside and San Bernardino counties. The sample has quotas for second and 1.5 generation groups (Mexicans, Vietnamese, Filipinos, Koreans, Chinese, and Central Americans from Guatemala and El Salvador) and includes three native-parentage comparison groups

comprised of third and later generation Mexican-Americans, Non-Hispanic Whites and Blacks.

- CILS involved a longitudinal survey of immigrant offspring living in San Diego and Miami (born abroad and raised in the United States or born in the U.S. to at least one foreign-born parent). The original survey was conducted in 1992, with samples of second-generation children attending the 8th and 9th grades in public and private schools in the metropolitan areas of Miami/Ft. Lauderdale in Florida and San Diego, California. The students were later sampled again as high school students in 1995-6, and finally as young adults in 2001-3. In total CILS surveyed 5,262 immigrant children in wave one, but only retained 3,334 respondents by wave three. Since we are interested in the final educational achievement we only use respondents from wave three of the data. CILS also asked in wave two about parental education and in addition directly interviewed the parents of approximately half of the original respondents. We use this additional information in supplementary analyses to assess the impact of error in childrens' reports of parental education on our estimates.

National Education Longitudinal Study (NELS): During the spring term of the 1987-1988 school year, the National Center for Education Statistics (NCES) initiated a nationally representative longitudinal study of 8th-grade students attending 1,052 high schools across the United States. A total of 24,599 8th-graders were surveyed in the base year of NELS:88. Subsamples of these cases were surveyed again for a total of 4 waves of interviews, the last one consisting of 12,144 cases interviewed in 2000, twelve years after the initial survey. At that time respondents were about 26 years old having finished secondary education. Since we want to assess overall educational achievement, we restrict our analysis to the fourth wave subsample.

To make our work comparable to previous research and to reduce issues of censoring, we use only respondents of an age when most will have finished their educational careers. For IIMMLA, ISGMNY and NELS, we restrict our samples to respondents ages 25 and above. As the CILS surveyed a younger population we have to use a lower cutoff and have a sample between ages 23 and 25. The second generation is

defined as children of at least one foreign born parent who were born in the US or arrived before starting primary school (less than 5 years old)⁵.

Variables:

National origin: For IIMMLA, CILS and the ISGMNY, we code national origin as respondent's place of birth. When the respondent was born in the United States, we use mother's place of birth; where this is missing or in the US, we use father's birthplace. NELS does not provide detailed information on parent's place of birth. For this survey, respondent's reported ethnic origin is used. All origin groups with at least 30 valid observations were used in the analysis.

Respondent's years of education: The data available on respondents years of education differed somewhat across the surveys: IIMMLA data contained greater detail on grade level and time spent in college, and provides a variable that maps this information into years of education, ranging from 0 to 20 years⁶. ISGMNY contains less detail about early schooling and thus we begin coding the lowest educational category (some grade school) at 6 years of education. The ISGMNY data is right truncated at 20 years of education. Similarly, CILS data was originally sampled in schools during early adolescence and thus is left censored at 10 years of schooling and the lack of information about time to degree right truncates the variable at 20 years of education for those with a professional or doctoral degree. NELS data is also left censored at 9 years of schooling and truncated at 20 years of schooling.

Parental years of education: In ISGMNY and IIMMLA, we coded parental education identically to respondent's and employ the same coding routine as described above. The IIMMLA included a small number of parents with no formal schooling, which we coded as zero years of education. CILS only provides categorical measures of educational attainment for respondent's parents, reducing the variation in parental education, with the lowest level at elementary school or less (coded as 6 years of education) and the highest

⁵ All analyses were replicated using a more restrictive definition of US born children of two foreign born parents and the results are essentially the same.

⁶ We restrict our analysis to those with 6 years of education or more, eliminating 6 observations.

level at college graduate or more (coded at 16 years of education). Similarly, NELS provides categorical measures of parental education that begin at “did not finish High School” (coded 10). The highest level of education, “Ph.D., M.D. etc” is coded as 20 years of education. We note that these estimates rely on student and parental translations of years of education and educational credentials from non-US educational institutions. Misreports of parental education may attenuate our estimates. However, we emphasize that the parental reports used in this paper are from less recently arrived immigrants with children who have gone through the US school system, and thus should have greater familiarity with US credentials. We also explore possible effects of child misreports on our estimates with further sensitivity testing below.

We coded both the number of years of formal education respondents’ mother and father received and then defined parental education as the highest of the two. Some of the analysis using aggregate data (e.g. Card 2000, 2005) uses only fathers education; with micro-data available however, we see no valid reason to assume that only fathers’ education is relevant. In any event our results are substantively robust to using only fathers or mothers education as the independent variable.

Analysis and Results

The analysis is straightforward. First we summarize and describe the distributions of parental and respondent’s education data for each national origin group sampled. We use means and variance as a measure of dispersion. We then calculate changes in these parameters between parents and children as well as the regression coefficient of respondents on parents’ education. This information is summarized in table 1.

	Parents			Respondents			Change across generations		Regression		
	mean	Var.	N	mean	Var	N	mean	Var	beta	Se	
IIMMLA: Mexican (MX)	9.7	18.3	399	13.3	5.7	462	3.5	-12.6	.13	.03	
Salvadoran (SV)	11.7	15.5	82	14.1	3.6	94	2.3	-11.8	.03	.05	
Guatemalan (GT)	11.4	17.4	52	14.0	3.9	58	2.6	-13.5	-.05	.06	
Chinese (CN)	15.2	18.0	145	16.5	3.2	152	1.2	-14.7	.14	.03	
Korean (KR)	15.3	9.8	162	16.1	4.0	169	0.8	-5.8	.05	.05	
Vietnamese (VN)	13.4	15.5	107	15.8	3.6	117	2.4	-11.9	.09	.05	
Filipino (PH)	15.7	5.0	187	15.5	3.4	189	-0.2	-1.6	.29	.06	
ISGMNY: China (CN)	13.1	19.2	86	16.0	2.8	90	2.9	-16.4	.13	.04	
Colombian (CO)	13.4	8.4	37	14.6	3.2	39	1.2	-5.2	.18	.09	
Dominican (DO)	11.6	13.3	114	14.2	5.2	122	2.5	-8.2	.09	.06	
Ecuadorian (EC)	12.5	11.6	57	13.8	4.0	59	1.3	-7.7	.19	.07	
Jamaican (JM)	13.9	4.5	32	14.9	3.3	34	1.0	-1.2	.05	.15	
Puerto Rican (PR)	11.7	8.8	137	13.1	5.6	146	1.4	-3.2	.01	.07	
CILS: Laotian (LA)	10.7	15.4	37	13.6	3.0	40	2.9	-12.5	-.07	.07	
Filipino (PH)	14.9	2.2	320	14.6	2.4	309	-0.3	0.2	.33	.06	
Thai (TH)	8.1	16.2	32	13.2	2.4	35	5.1	-13.8	.03	.07	
Vietnamese (VN)	11.4	12.7	88	14.8	3.2	88	3.5	-9.5	.10	.06	
Mexican (MX)	10.4	12.3	259	13.4	3.0	251	3.0	-9.3	.05	.03	
Cuban (CU)	13.2	6.5	524	14.4	3.5	521	1.2	-3.0	.20	.03	
Dominican (DO)	12.9	11.2	37	13.7	3.5	37	0.7	-7.7	.18	.09	
Haitian (HT)	12.1	10.8	46	14.5	1.8	48	2.4	-9.0	-.03	.06	
Jamaica (JM)	14.3	4.7	44	15.2	3.4	46	0.9	-1.3	.22	.12	
Nicaraguan (NI)	13.9	8.3	74	14.3	3.2	73	0.4	-5.1	.14	.07	
Colombian (CO)	13.5	4.5	71	14.3	3.4	72	0.8	-1.2	.33	.10	
NELS: Mexican (MX)	11.6	4.6	326	13.3	2.1	375	1.8	-2.5	.10	.04	
Cuban (CU)	15.0	9.0	35	14.6	3.8	39	-0.3	-5.3	.14	.11	
Puerto Rico (PR)	12.9	7.3	73	13.4	1.7	79	0.5	-5.6	.06	.06	
Indian (IN)	18.2	4.2	48	16.3	2.4	50	-1.9	-1.9	.30	.10	
Chinese (CN)	15.6	11.6	96	15.6	2.4	108	0.0	-9.1	.08	.05	
Filipino (PH)	15.6	6.3	49	15.0	4.1	55	-0.6	-2.2	.40	.11	
Korean (KR)	16.1	8.7	60	15.3	2.9	65	-0.8	-5.9	.23	.07	
<i>Comparison groups with native parents</i>											
NH White (NHW): NELS	14.3	7.4	6477	14.2	3.4		6797	-0.2	-4.0	.29	.01
ISGMNY	15.3	8.2	246	15.5	5.8		249	0.1	-2.4	.41	.05
IIMMLA	15.0	7.6	309	14.8	5.2		318	-0.2	-2.4	.31	.04
NH Black (BLK): ISGMNY	13.7	6.9	229	13.5	4.6		249	-0.2	-2.3	.28	.05
NELS	13.7	6.3	714	13.6	2.8		779	0.0	-3.6	.20	.02
IIMMLA	14.0	6.5	294	13.8	3.7		312	-0.1	-2.7	.22	.04
Mexican 3rd gen. (MX3): IIMMLA	12.9	4.9	271	13.4	4.2		289	0.5	-0.7	.38	.05
Puerto Rican (PR): ISGMNY	12.4	5.9	50	13.2	5.5		53	0.7	-0.4	.12	.14

Table 1: Educational achievement as measured in years of respondents, their parents and intergenerational change by national origin group and survey source in our analysis. Coefficients that are significant at the 0.05 level or higher are in bold.

We see that with the exception of Filipinos the children of immigrants have a higher average education than their parents and that the distribution of education is far more compressed. Thus not only do the children of immigrants have higher educational attainment than their parents, the variance of the distribution is much lower in the second-generation.

The last two columns show the estimated slope coefficients and standard errors from a linear regression of respondents education on parental education measured in years. The slopes vary significantly across groups ranging from being statistically not different from 0 to a maximum of 0.41. Thus a key assumption of ecological regression - that the relationship between dependent and independent variables is equal across groups - is not satisfied. More specifically, we see that among immigrant groups, with the exception of Filipinos, Colombians and Indians, the effect of parental education is substantially smaller than for Whites with native parents where the regression coefficient ranges from 0.29 (NELS), 0.31 (IIMMLA) to as high as 0.41 in the ISGMNY data – coefficients of the same magnitude as the ones quoted by Card (2000, 2005). Among native Blacks the influence of parental education is similar, 0.28 in the ISGMNY data but a bit lower in the NELS and IIMMLA data. The censoring of parental education at 10 years likely contributes to these lower coefficient estimates in the NELS data. Only among the children of native born Puerto Ricans is parental influence as low as that observed among the children of immigrants.

In our other third generation immigrant group, families with Mexican ancestry in the Los Angeles area, the effect of parental education on respondent's education is with 0.36 significantly higher than the coefficients observed amongst the second generation groups. This suggests that immigrant status is a decisive factor in increasing educational intergenerational mobility amongst those with Mexican origins.

To replicate results from previous analysis, we average years of education for the second generation respondents and their highest educated parent for each origin group

and then use the aggregate data to regress group averages of respondents' education on parental education. Model 1 in table 2 presents the results of this analysis. The slope coefficient of this regression is 0.34 and once we weight each national origin group to represent their proportion of the US foreign born population as of 2000 (model 1b) we obtain a regression coefficient of about 0.43. These are coefficients of the same magnitude as the one found by Card et al (2000) and Card (2005) using the same methodology. It is also significantly higher than *almost all the* coefficients of the regressions that estimate intergenerational transmission within groups in table 1.

Model 2 uses the pooled individual level data for all immigrant groups from all our surveys, weighted for their representation among the US foreign born, thus giving an average of the effect of parental education in immigrant families. This model also includes dummy variables for each national origin to net out differences in average education levels of groups. The slope estimate for the effect of the transmission of parental educational achievement is 0.11, significantly lower than the 0.3 to 0.4 estimated for non-migrant families.⁷

Finally in Model 3 we enter both individual information on parental education as well as the mean parental education of each group – in effect estimating equation 1 from above and disaggregating the individual and group level effects. As expected, the effect of average education in the group β_2 is not zero (or negligible) as would be required for reliable inference with group level data, but in fact is larger in magnitude by a factor of about 2 as compared to the effect of parental educational achievement. Taken together these two coefficients add up to the aggregate level estimate obtained in model 1.

⁷ A regression coefficient obtained without weights is an almost identical 0.12.

	Model 1			Model 1b			Model 2			Model 3			Model 4		
	Coef.	se	t	Coef.	se	t	Coef.	se	t	Coef.	se	t	Coef.	se	z
Intercept	10.16	0.72	14.13	8.86	0.73	12.13	12.02	0.43	27.98	11.85	0.1	117.89	10.28	0.76	13.52
Mean of Parental Education	0.34	0.06	6.05	0.43	0.06	7.46							0.21	0.06	3.53
Parental education							0.11	0.01	12.66	0.15	0.008	18.81	0.12	0.01	13.34
National origin index							yes			no					
Weighted to US population	no			yes			yes			yes			no		
N	19			19			4038			4038			4038		

Table 2: Models estimating intergenerational transmission of education in immigrant families pooling IIMMLA, ISGMNY, CILS and NELS data.

Measurement Error

One possible objection to this analysis may be the issue of measurement error. As discussed by Borjas (1992), measurement error in parental education (due to recall error for example) may increase the estimate of the effect of mean education of the group β_2 . Acting as an instrument of sorts the mean parental education may capture some of the individual level effects that are “lost” due to measurement error. However, as Borjas (1995) later shows using multiple measures of parental skills, the magnitude of this effect is not substantial enough to significantly alter the results, especially in the case of education where measurement error seems more limited.

Using a subset of our data – the CILS - we directly address this issue. The CILS asked second-generation respondents in wave 2 and wave 3 about parental education and included a parental questionnaire for a subset of the sample. While responses are highly correlated (.77) they are far from identical, pointing to some measurement error. To assess to what extent this measurement error may attenuate the coefficients for intergenerational transmission and inflate the estimated magnitude of “group effects” we used a latent variable model with all three measures of parental education as indicators of a latent variable that is then included in the regression equation for educational outcomes along with a vector of the origin group means. Thus we take the “true” educational achievement of a parent as a latent variable η_i that is in turn measured by a vector of observed indicators \mathbf{x}_i . In our case \mathbf{x}_i has length three combining the respondents answers about their parents education in wave 2 and wave 3 as well as the parental questionnaire where available. Vectors of factor loadings λ and intercepts τ relate these measured indicators to our unmeasured variable parental education leaving a vector of normally distributed residuals ζ_i . This measurement model can be written as:

$$\mathbf{x}_i = \boldsymbol{\tau} + \boldsymbol{\lambda}\boldsymbol{\eta}_i + \boldsymbol{\zeta}_i \quad (3)$$

In conjunction with equation 1 this gives us a regression coefficient for family level transmission of education that is not attenuated by measurement error.⁸ We estimate this model using a full information maximum likelihood estimator as implemented in M-plus (Muthén and Muthén 2007).

Table 3 summarizes the results of this endeavor. For comparison we include regression models analog to those in model 3 from table 3. Using measures of fathers' education we see that the latent variable estimate of parental transmission is indeed somewhat higher as compared to the regression estimates while the effect of national origin education is a bit lower. In the case of mothers education we see a similar pattern in the estimate of the parental transmission but the "group effect" does not reach statistical significance in either the regression or the latent variable models. We conclude that measurement error indeed does introduce some upward bias on the estimated effect of characteristics of the national origin groups and some downward bias on the estimate of family level transmission. However, the magnitude of this bias is not large enough to substantively alter the conclusions of our analysis.

	Regression					
	Using Wave I		Using Wave II		Latent Variable Model	
	Coef.	z	Coef.	z	Coef.	z
Fathers Education	0.09	5.78	0.09	5.47	0.12	4.32
Mean of fathers educ.	0.20	1.89	0.16	1.67	0.14	1.81
N	1312		1289		1559	
Mothers Education	0.13	8.06	0.11	7.00	0.16	4.18
Mean of mothers educ.	0.12	1.61	0.09	1.31	0.04	0.42
N	1383		1340		1559	

Table 3: Models using various different measurements of parental education available in the CILS data. The regression models are estimated analog to model 3 in table 3. All standard errors are adjusted for clustering. The latent variable model is estimated using a robust maximum likelihood estimator and shows excellent fit to the data: CFI>0.99, RMSEA <0.05.

⁸ As a measure of group mean education we take the average of wave 2 responses. The correlation of national origin group means is with 0.96 (fathers) and 0.94 (mothers) very high.

Finally we want to briefly address two other caveats to our analysis. First, the majority of our data, the three second-generation surveys, come from large cities with large numbers of immigrants and where a disproportionate number of migrants live in ethnic neighborhoods. Our paper therefore best represents the experiences of immigrants and their children in traditional gateway cities. However, this representation is valid for the majority of the immigrants in the United States: according to the US Census in 2010, 38% of immigrants lived in New York, Los Angeles, Miami, Chicago, and Houston alone, and 85% of immigrants lived in the 100 largest metro areas of the US. To further assess whether a national-level sample would differ, we replicated all the results above using only the NELS national level data, and applying NELS survey weights for national representativeness using the Stata 12 subpopulation commands. The substantive finding remained the same: the effect of average group education level (0.33, for 10 groups) was much larger than the effect of individual level parental education (.19), although both were larger in magnitude than the sample used in this paper.

Second, there is some discrepancy in the characteristics of immigrant national origin groups across surveys: for instance, estimates of intergenerational transmission among Mexicans in the IIMMLA and NELS survey are higher, and statistically significant, whereas estimates from CILS data are lower and not statistically significantly different from 0. There are many differences between each survey that could account for these differences: sampling at different age points (youth in NELS and CILS, and adults in IIMMLA and ISGMNY), sampling metropolitan areas instead of nationally, the slightly different age ranges, especially the younger age of CILS and NELS respondents. Another possible culprit is the censoring and truncation in our education variables; however, there seems to be no consistent upward or downward patterns between the surveys that share immigrant national groups, despite survey level differences in the educational coding. Ultimately we cannot pin down the cause for these differences – in our case we take comfort in the fact that estimates are *substantively* consistent – for instance, that Filipinos consistently have the highest levels of intergenerational transmission whereas most of the groups show estimates that are below 0.2. More generally this variance in estimates should remind us that analyses from just one survey,

even when the survey is of high quality, should be interpreted with extreme caution as they may not be representative of the larger phenomenon.

Discussion:

This paper has shown that inference about intergenerational mobility in migrant families drawn from group level data are not comparable to estimates obtained from regressions that rely on individual level, parent-child dyad information. The former contain both the effect of parental education and the significant effects of group level educational characteristics and associated variables of the national origin groups.

In themselves of course neither the individual nor aggregate level approaches are “wrong” or “right” - rather they answer different questions. If we want to know how the immigrant – national origin or ethnic - *groups* will fare across generations, then a method that includes family level as well as group level factors is acceptable. As Borjas correctly points out and we confirm in this analysis, the group level effects are significant in the case of immigrants – about twice the size in magnitude as compared to family level transmission. On the other hand if we are interested to what extent the educational intergenerational mobility of immigrants compares across time, or to the intergenerational mobility in the native population, then only data that allows us to link parent-child dyads will give the correct answer.

The distinction also speaks to different understandings of assimilation – at what level do relevant social processes occur and what are the constituent social elements in the theory. If we take social or ethnic *groups* as the constituent elements of society as early Chicago School theories – who studied social relations between ethnic groups – did, then analysis based on aggregate data that combines family and group level processes in one single estimate is perfectly acceptable. However, contemporary social science theories of assimilation which focus on socio-economic mobility – most prominently the rational choice based neo-classic assimilation model of Alba and Nee (1997; 2003) - have abandoned this group-based approach and take assimilation chiefly as an individual level process. A process where ethnicity and group level processes certainly play a role but

ethnic groups are neither the building blocks of society nor the units of analysis (see also Brubaker 2004; Wimmer 2009). In this case distinguishing between individual level and aggregate level processes is essential.

In the case of educational achievement among immigrants, group-level effects and family transmission add up to a coefficient of about the same magnitude than the pure inter-family transmission in native families. Yet the family level transmission component is much lower in immigrant families as compared to natives. Group level mechanisms that are specific to immigrants, such as discrimination or ethnic social capital, are certainly part of the explanation for this difference. Also in the case of migrants formal education may not be a reliable signal for human capital, especially among those from countries with unequal access to education or poorly functioning education systems. Thus the issue of “measurement error” is also a larger conceptual point. When what we are really interested in is the human capital of migrant families and its effect on the social reproduction of inequity, then in the case of migrants’ formal education may not be the best variable to assess it.

Also when thinking about the long-term implications of immigration on social stratification, the difference between family level and group level is pertinent. Comparisons across time that state a consistent rate of intergenerational mobility but are based on aggregate level analysis may miss shifts in the relative importance of group level versus family level factors. Finally, when thinking in policy terms about intergenerational mobility, this analysis suggests that immigrant formal education per se may not be the most important predictor of the educational outcomes of their offspring. The exact conclusions of course will depend on the nature of these group level effects, whether they are due to discrimination against certain groups, differences in the ethnic social capital or some other process.

References

- Alba, Richard D and Victor Nee. 1997. "Rethinking Assimilation Theory for a New Era of Immigration." *International Migration Review* 31:826--874.
- Alba, Richard D. and Victor Nee. 2003. *Remaking the American Mainstream : Assimilation and Contemporary Immigration*. Cambridge, MA: Harvard University Press.
- Aydemir, Abdurrahman, Wen-Hao Chen, and Miles Corak. 2008. "Intergenerational education mobility among the children of Canadian immigrants." *IZA Discussion Paper No. 3759*.
- Bankston, Carl L. and Min Zhou. 1995. "Religious participation, ethnic identification, and adaptation of Vietnamese adolescents in an immigrant community." *The Sociological Quarterly* 36:523--534.
- Bauer, Philipp C. and Regina T. Riphahn. 2006. "Timing of school tracking as a determinant of intergenerational transmission of education." *Economics Letters* 91:90-97.
- Blau, Peter M. and Otis D. Duncan. 1967. *The American occupational structure* New York Wiley.
- Borjas, George J. 1992. "Ethnic Capital and Intergenerational Mobility." *Quarterly Journal of Economics* 107:123-150.
- . 1993. "The Intergenerational Mobility of Immigrants." *Journal of Labor Economics* 11:113-135.
- . 1994. "Long-Run Convergence of Ethnic Skill Differentials - the Children and Grandchildren of the Great Migration." *Industrial & Labor Relations Review* 47:553-573.
- . 2006. "Making it in America: Social mobility in the immigrant population." *Future of Children* 16:55-71.
- Brubaker, Rogers. 2004. "Ethnicity without groups." Pp. 7-27 in *Ethnicity without groups*. Cambridge, Mass.: Harvard University Press.
- Card, David. 2005. "Is the new immigration really so bad?" *The Economic Journal* 115:300--323.
- Card, David, John DiNardo, and Eugena Estes. 1998. "The More Things Change: Immigrants and the Children of Immigrants in the 1940s, the 1970s, and the 1990s." *NBER Working Paper 6519*.

- . 2000. "The More Things Change: Immigrants and the Children of Immigrants in the 1940s, the 1970s, and the 1990s." Pp. 227-270 in *Issues in the Economics of Immigration*, edited by G. J. Borjas. Chicago, IL: University of Chicago Press.
- Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. 2008. "When are ghettos bad? Lessons from immigrant segregation in the United States." *Journal of Urban Economics* 63:759-774.
- d'Addio, Anna Cristina. 2007. "Intergenerational Transmission of Disadvantage: Mobility or Immobility across Generations? A Review of the Evidence for OECD Countries." *OECD Social, Employment and Migration Working Papers*.
- Dustmann, Christian. 2008. "Return migration, investment in children, and intergenerational mobility - Comparing sons of foreign- and native-born fathers." *Journal of Human Resources* 43:299-324.
- Dustmann, Christian and Albrecht Glitz. 2011. "Migration and Education." Norface Research Programme on Migration, Department of Economics, University College London.
- Firebaugh, Glenn. 1978. "A Rule for Inferring Individual-Level Relationships from Aggregate Data." *American Sociological Review* 43:557-572.
- Gang, Ira N. and Klaus F. Zimmermann. 2000. "Is Child like Parent? Educational Attainment and Ethnic Origin." *Journal of Human Resources* 35:550-569.
- Gibson, Margaret A. 1988. *Accommodation without Assimilation: Sikh immigrants in an American high school*: Cornell University Press.
- Goodman, Leo A. 1953. "Ecological Regressions and Behavior of Individuals." *American Sociological Review* 18:pp. 663-664.
- Gordon, Milton M. 1964. *Assimilation in American Life: The Role of Race, Religion, and National Origin*. New York: Oxford University Press.
- Hammond, John L. 1973. "Two sources of error in ecological correlations." *American Sociological Review* 38:764--777.
- Hout, Michael and Thomas A. DiPrete. 2006. "What we have learned: RC28's contributions to knowledge about social stratification." *Research in Social Stratification and Mobility* 24:1-20.
- Jargowsky, Paul A. 2004. "The ecological fallacy." *The Encyclopedia of Social Measurement* 1:715--722.

- Kasinitz, Philip, John H. Mollenkopf, Mary C. Waters, and Jennifer Holdaway. 2008. *Inheriting the City: The children of immigrants come of age*: Harvard University Press.
- Mare, Robert D. 1981. "Change and stability in educational stratification." *American Sociological Review*:72-87.
- Massey, Douglas S. 1998. *Worlds in Motion: Understanding International Migration at the End of the Millennium*: Oxford University Press.
- Muthén, Linda K. and Bengt O. Muthén. 2007. *Mplus User's Guide. Fifth Edition*. Los Angeles, CA.
- Nagel, Joane. 1995. "American Indian Ethnic Renewal: Politics and the Resurgence of Identity." *American Sociological Review* 60:947--965.
- Nielsen, Helena S., Michael Rosholt, Nina Smith, and Leif Husted. 2003. "The school-to-work transition of 2 nd generation immigrants in Denmark." *Journal of Population Economics* 16:755-786.
- Park, Julie and Dowell Myers. 2010. "Intergenerational mobility in the post-1965 immigration era: Estimates by an immigrant generation cohort method." *Demography* 47:369-392.
- Park, Robert E. . 1930. "Assimilation, social." Pp. 281–83. in *Encyclopedia of the Social Sciences*, edited by E. Seligman and A. Johnson. New York: Macmillan.
- Piore, Michael J. 1980. *Birds of passage : migrant labor and industrial societies*. Cambridge; New York: Cambridge University Press.
- Portes, Alejandro and Rubén G Rumbaut. 2001. *Legacies : the story of the immigrant second generation*. Berkeley: University of California Press.
- Riphahn, Regina T. 2003. "Cohort effects in the educational attainment of second generation immigrants in Germany: An analysis of census data." *Journal of Population Economics* 16:711-737.
- Robinson, William S. 1950. "Ecological Correlations and the Behavior of Individuals." *American Sociological Review* 15:351--357.
- Smith, James P. 2003. "Assimilation across the Latino Generations." *American Economic Review* 93:315-319.
- Tilly, Charles. 1998. *Durable inequality*. Berkeley, Calif.: University of California Press.

- Waldinger, Roger D. and Michael I. Lichter. 2003. *How the other half works : immigration and the social organization of labor*. Berkeley, Calif.: University of California Press.
- Warner, William Lloyd and Leo Srole. 1945. *The Social Systems of American Ethnic Groups*. New Haven, CT: Yale University Press
- Waters, Mary C. 1999. *Black identities: West Indian immigrant dreams and American realities*. New York: Russell Sage Foundation.
- Wimmer, Andreas. 2009. "Herder's heritage and the boundary-making approach: Studying ethnicity in immigrant societies." *Sociological Theory* 27:244--270.