Does 'Ethnic Capital' Matter? Identifying Peer Effects in the Intergenerational Transmission of Ethnic Differentials*

Alexis León°

August 2005

Abstract

An instrumental variables strategy is employed to determine whether the observed association between individual human capital and average skills in ethnic groups (ethnic capital), even after conditioning on parental skills, is due to ethnic peer effects. The instrument for ethnic capital is derived from the occupational mix of US immigrants arrived in the 1900's and 1910's, while father's age at arrival instruments for parental skills. Using US Census data on adult literacy in English and children's school attendance, I find evidence of a persistent ethnic capital effect. High geographic concentration and high endogamy rates tend to accentuate this effect.

Keywords: human capital accumulation, peer effects, immigration, ethnicity, intergenerational mobility

JEL Classification: J15, J24, J62

---

* I owe special thanks to Joshua Angrist and Daron Acemoglu for their valuable advice, guidance and encouragement throughout this project. I also thank David Autor, Lucia Breierova, Amitabh Chandra, Ira Gang, María Guadalupe, Robert LaLonde, David Lyle, Gerard Padró-i-Miquel, Michele Pellizzari, Michael Piore, Hans-Joachim Voth, Pierre Yared, and seminar participants at MIT, the University of Pittsburgh, the University of Chicago, the Humboldt University in Berlin, the Bank of Spain, and the Immigration Conference at the Federal Reserve Bank in Philadelphia. All remaining errors are my own.

° Department of Economics, University of Pittsburgh, 4530 Wesley W. Posvar Hall, Pittsburgh, PA 15260. E-mail: aleon@pitt.edu.

I. Introduction

Differences in socioeconomic outcomes across ethnic groups tend to persist over time. A significant part of the inter-group variation in measures of human capital or labor market performance is transmitted across generations, thus slowing down the process of ethnic convergence that could be expected from simple mean reversion. Table 1 illustrates how much of the ethnic differences in education that existed among US immigrants in 1910 persisted into the second generation, for several national origin groups. For instance, Scottish male immigrants aged 30 to 50 were 35.6 percentage points more likely to be literate than Italian immigrants in 1910, and those in turn were 17.4 points more likely to be literate than Mexican immigrants. School enrollment rates of second-generation Scots aged 6 to 18 in 1910 were 8.2 percentage points above those of Italians, which were in turn 28.4 points above the attendance rate for Mexicans. Thirty years later, there existed substantial differences in educational attainment among second-generation adults in those same groups: average years of schooling in 1940 were 10.1 for Scottish-Americans, 8.7 for Italian-Americans and only 4.1 for Mexican-Americans. Their children also had noticeably different school attendance rates: a third-generation Scottish-American of schooling age was 6 percentage points more likely to be enrolled in school than a third-generation Italian-American child, who was, in turn, 12.1 points more likely to be attending school than a third generation Mexican-American.

The persistence in ethnic differentials over time could simply be the result of the transmission of skills that takes place within the family. Parents can influence the socioeconomic development of their offspring by investing time, effort, and financial resources in their children's human capital.[1] Other individuals in their ethnic group, however, can influence children as well. Friends, relatives, and neighbors can also serve as role models, spend time helping with school homework, and transmit certain attitudes towards education and work. Hence, being exposed to an advantageous ethnic environment while growing up can also contribute to the children's human capital

[1] While acknowledging that the genetic transmission of ability can also be an important channel, it is not the subject of interest in this paper to determine whether the passage of parental skills to children can mostly be attributed to nurture or not.

accumulation process, beyond the direct role of their parents. The existence of peer effects in the ethnic group will then exacerbate the extent to which the skill level in the immigrant generation determines the socioeconomic success of the next generation. That, in turn, will have implications for overall inequality in the economy.

Following the predominant terminology in the literature, as introduced in Borjas (1992), I will refer to these ethnic peer effects as the 'ethnic capital' effect in the intergenerational transmission of skills. 'Ethnic capital' denotes the average in the ethnic group of some measure of skills or socioeconomic performance –as opposed to 'parental capital', which designates the corresponding measure for a given individual's parents. While this ethnic spillover may operate primarily through geographic concentrations of peers in the same ethnic group, ethnic capital effects are not to be confused with local (or 'neighborhood') effects. Even within a neighborhood, children are more likely to befriend and interact with other individuals in the same ethnic group,[2] in which case the impact of peers of the same ethnicity will outweigh that of neighbors in other groups. Similarly, relatives or friends of the same ethnic background who do not live in the immediate neighborhood can serve as role models and exert an influence on the child, thus contributing to the diffusion of ethnic capital.[3]

The main challenge in disentangling the two channels of intergenerational transmission of skills, and therefore estimating the parental and ethnic capital effects separately, is identification. Despite the potential importance of this question for immigration and welfare policy, most studies to date have relied primarily on ordinary least squares (OLS) regression strategies to study ethnic spillovers in the transmission of skills across generations. As I argue below, however, parental skills and average skill levels in the ethnic group may be correlated for a number of reasons, so an observed association between ethnic capital and child outcomes is not necessarily causal. To solve this problem, I use instrumental variables (IV) to estimate both parental and ethnic capital effects consistently.

[2] Alba (1992) showed this for second-, third- and higher generation children in several Caucasian European-American ethnic groups.

[3] Since residential segregation is one of the main channels through which ethnic capital is transmitted across cohorts within groups, my primary interest is not to isolate pure peer effects from local (neighborhood) effects. See Section IV, however, for suggestive evidence on the role of geographic concentration in the magnitude of the ethnic capital effect.

My identification strategy exploits variations in the occupational mix of new immigrant flows over time and across ethnic groups, much in the same way as Angrist (2002) did for the effect of changing sex ratios. The instruments for average ethnic skills are derived from official records of immigrant arrivals by year and national origin, further classified into broad occupation groups. Variation in the occupational mix of new immigrants is assumed to be exogenous to changes in local economic conditions in the US, and therefore solves the endogeneity problem caused by ethnicity-year specific shocks encouraging further skill accumulation. The inclusion of ethnicity and year main effects and of individual characteristics as control variables ensures that my results are robust to group-specific characteristics that might be correlated with transferable skills of fathers and with the occupational mix of new immigrants, such as tastes for work or education. This strategy would fail, however, in the presence of time-ethnicity specific shocks that affected both transferable skills of fathers and occupational mix of recent arrivals *and* that were not fully captured by the covariates in my regressions. Such a situation appears unlikely, though, particularly since the results appear robust to the inclusion of additional control variables and to several other specification checks, thus weighing in favor of a causal interpretation.

I also instrument for a second key endogenous regressor, parental skills (measured by father's literacy), with father's age at arrival interacted with a dummy variable for non-English speaking country of birth, as in Bleakley and Chin (2003). The inclusion of a father's age-at-arrival main effect controls for additional (non-literacy related) unobserved dimensions of skills that may be transmitted from parents to children.

In order to clarify this idea, consider, for example, the children of Italian-American immigrants in the US at the beginning of the 20th century. My strategy relates changes in school attendance rates of second-generation Italian-American children between 1910 and 1920 to changes in the fraction of recently arrived Italian immigrants who were recorded as having low-skilled occupations (agricultural workers, laborers and servants). The 'experiment' behind this approach consists, then, in observing how distinct communities of immigrants will be affected by the arrival of newcomers with a different level of skills. In fact, there is anecdotal evidence of incumbent immigrant groups at various points in time being alarmed by the arrival of what they perceived to be 'lower

quality' immigrants in their ethnic groups.[4] The existence of ethnic capital effects would provide some basis for the fear that new low-skilled waves of immigrants could dilute the skills of the community and have a negative impact on the next generation as well.

This strategy constitutes a good natural experiment because, as I will argue below, the resulting variation in the average skills by ethnicity was driven mainly by home country conditions in the early twentieth century (most notably World War I), which were exogenous to local US market conditions facing the existing immigrants and their children. Moreover, social interactions among individuals within each of the ethnic groups used in this analysis were indeed more important than with individuals outside the group, as evidenced by the high intra-group marriage rates that will be presented below. Finally, the immigrants (and immigrant flows) studied in this analysis constitute a major demographic episode in American history, with aliens arriving in numbers that went unmatched for almost a century.[5]

An additional contribution of this paper is the use of measures of ethnic capital that are contemporaneous with child outcomes (as opposed to using skills of immigrants in a previous period), to better reflect the actual environment facing children and reduce the potential bias from return migration in the measure of average skills in the ethnic group. Another improvement is the use of repeated cross sections, which allows me to control for ethnicity and year main effects.

[4] For instance, Thomas Sowell (1981, pp. 107-108) notes that "the relationship between the earlier arriving members of a group and those arriving later is an important factor in the history of most American ethnic groups. (…) The earlier Italian immigrants had gained a measure of acceptance and prosperity by the time the massive waves of southern Italians arrived. (…) The northern Italians openly repudiated the southern Italians. Many even preferred to pass for Americans." Irving Howe (1976, p229) remarks that "by the turn of the [20th] century, the tensions between the established German Jews and the insecure East European Jews had become severe –indeed, rather nasty. (…) The Germans found it hard to understand what could better serve their ill-mannered cousins than rapid lessons in civics, English, and the uses of soap." In both cases, however, the newcomers did interact with the existing communities, as evidenced by the high marriage rates within each group. Common culture, language or history could help explain why, for example, "German Jews established and financed schools, libraries, hospitals, and community centers to aid, and especially to Americanize, the eastern European Jewish immigrants." (Sowell, p.81).

[5] Borjas (1994) refers to the huge flow of immigrants between 1880 and 1924 as the *First Great Migration*, to distinguish it from the *Second Great Migration* that took place in the last twenty years of the twentieth century: the number of immigrants admitted to the United States in the decade 1901-1910 is recorded at 8.8 million (Ferenczi and Willcox (1929)), or 10.4% of the population, which was only exceeded nine decades later with the arrival of more than 9 million legal immigrants, or 3.4% of the population, between 1991 and 2000.

The variables of interest in this research are human capital outcomes such as a proxy for literacy in English for adults and school attendance for children. Both are relevant education measures in the period being studied. Using micro data from the 1910 and 1920 US Censuses, I find evidence of significant ethnic capital effects in the intergenerational transmission of skills. The IV estimates are slightly, though not significantly, lower than the OLS estimates, which are subject to omitted variables bias and attenuation bias. IV estimates of the direct parental effect are much higher than the OLS estimates, suggesting severe measurement error in father's skills (the literacy variable). The results also suggest that ethnic spillovers are stronger where the geographic concentration of immigrants is highest. This result is consistent with ethnic peer effects that operate, at least in part, through neighborhood effects. Finally, regressions that take into account differences in endogamy rates by region also indicate that peer effects are larger for more endogamous communities, while insignificant for ethnic groups in regions where endogamy is very low.

The rest of the paper is organized as follows. Section II presents a theoretical model of ethnic peer effects, develops the estimation framework, and highlights the econometric issues involved in attempting to disentangle parental from ethnic peer effects in the intergenerational transmission of skills. Section III describes the data and presents the base empirical results. Section IV discusses some robustness checks and additional results. Section V summarizes the paper and concludes.

II. Background

A. Theoretical Framework

The idea that ethnic skills are transmitted across generations can be rationalized by Borjas' (1992) 'ethnic capital' model, a theory of human capital externalities. Similar ideas appear in the sociology literature on 'social capital:' Loury (1977) first introduced this term to explain how race differences in earnings persist over time due to spillover effects within a racial group; Coleman (1988) later developed that concept and applied it to the study of peer effects in the academic performance of high school students. More

recently, Putnam (1995, 2000) introduced the notion of 'social capital' in the political discussion of the decreasing participation in civic organizations in the US.

In this framework, utility-maximizing parents invest in the human capital of their children, while ethnic human capital has an external effect on the production of children's human capital.[6] Denoting parental human capital stock by h_t and children's human capital stock by h_{t+1}, this last assumption can be formalized as the following production function:

$$h_{t+1} = \theta\left(s_t h_t\right)^{\alpha_1} \bar{h}_t^{\alpha_2}, \tag{1}$$

where s_t is the fraction of h_t devoted to the production of h_{t+1} (and hence not used to produce, or be exchanged for, consumption goods), and $\bar{h}_t$ is the average human capital in the parents' peer group. The model is, therefore, characterized by dynamic externalities, in the sense that the human capital of one generation contributes to the production of the next generation's human capital. This is meant to capture the influence that other adults in the ethnic group outside the immediate family have on the education of infants and teens. Other things equal, interaction with peers and exposure to the cultural norms and values (and to examples of rewarding work and achievement) that are characteristic of a particular ethnic group affect children's human capital accumulation process. A more recent application of this idea can be found in Lundberg and Startz (1998), who use a similar model to explain the persistence of the racial wage gap.

As derived in Borjas (1992), the utility-maximization problem is solved by a function f determining the parental supply of time allocated to investing in children's human capital:

$$s_t = f\left(h_{t+1}, \bar{h}_t\right). \tag{2}$$

Substituting (2) into (1) yields the level of human capital of the next generation. I take a logarithmic transformation to obtain:[7]

$$\log h_{t+1} = \beta_0 + \beta_1 \log h_t + \beta_2 \log \bar{h}_t \tag{3}$$

---

[6] In Loury's terms, the opportunities of young people to acquire skills depend both on "the quality of home environment" as well as "the quality of the community environment." (Loury, 1977, p.159).

[7] It can also be shown that $\partial \log \bar{h}_{t+1} / \partial \log \bar{h}_t < 1$ iff $\alpha_1 + \alpha_2 < 1$ (condition for convergence of ethnic skill differentials). For further discussion of this and other implications of the ethnic capital model, see Section II in Borjas (1992).

The parameter β_2, the coefficient on average human capital, represents a peer group effect. This concept has been an object of considerable interest among economists: Benabou (1993) analyzes how residential segregation concentrates low-skilled learners in schools, which affects the learning process and results in persistent and widening income inequality. Bertrand, Luttmer and Mullainathan (2000) investigate network effects in welfare participation; Sacerdote (2001), Zimmerman (2003), and Winston and Zimmerman (2004) study peer effects in academic outcomes among college roommates; while Hoxby (2000), and Angrist and Lang (2004) estimate peer effects in the classroom.[8] They do not always, however, provide convincing evidence. The difficulty of controlling for confounding factors, even in those settings where potential problems of endogenous sorting into groups are resolved through randomized assignment, casts some doubts on the identification strategy of many of those studies.

In practice, of course, child outcomes are determined by many other factors beside parental skills and peer effects. I therefore add a stochastic error term to (3), as well as a vector of individual covariates z_i that includes region effects, age, father's age, and other demographic variables. Also, I adopt notation that reflects that (i) different individuals belong to different ethnic groups and that the externality will take place at the ethnic peer group level, and (ii) individuals in my data are observed in different years. The resulting equation is:

$$y_{ijt} = \beta_0 + \beta_1 x_{ijt} + \beta_2 \bar{x}_{jt} + \delta_j + \delta_t + z_{it}^{'}\gamma + \varepsilon_{ijt}, \qquad (4)$$

where y_{ijt} is an observable socio-economic outcome of child i in ethnic group j at time t (such as school enrollment), x_{ijt} is a measure of skills of the father (of child i in ethnicity j at time t), $\bar{x}_{jt}$ is the average skills of individuals in the father's generation in ethnic group j at time t, δ_j and δ_t are ethnicity and Census year effects, respectively, and ε_{ijt} is an individual error component.[9]

[8] To borrow Manski's (1993) terminology, the intergenerational transmission parameter β_2 in equation (1-4) expresses a "contextual or exogenous effect," as opposed to an "endogenous social effect," which is the case of the peer effects studied by Zimmerman, Sacerdote, or Hoxby.

[9] This "linear in group means" specification, which results from the assumption that the average human capital of the ethnic group enters the production function (1), may in general not be the only or the best way to capture peer group effects. Other relevant measures of the distribution of skills in the peer group, such as the variance, the maximum or the minimum, could potentially have an impact as well. In this study, however, the use of a binary variable (literacy) as the skill measure makes exploring such alternatives

B. Econometric Framework

The most important identification problem raised by equation (4) is omitted variable bias from correlation between average skills in the ethnic group, $\bar{x}_{jt}$, and other ethnicity-year effects contained in the residual term ε_{ijt}. For example, if different ethnic groups are not distributed proportionally across occupations, industries or geographic areas, then an economic shock that increases opportunities relatively more for a given ethnic group at some point in time will encourage accumulation of skills for adults in that group, while at the same time increase the schooling of their children. Moreover, $\bar{x}_{jt}$ is subject to measurement error and is affected by economically motivated return migration, as well as by immigrant arrivals.

In order to solve these problems, the fraction of 'low skilled' new immigrants to which different ethnic groups in different years were exposed is used to construct an instrument for $\bar{x}_{jt}$. Recent flows of unskilled immigrants are correlated with the average skills in the ethnic group because those new immigrants (arrived in the 5 years prior to the Census year) are least likely to have returned to their home countries, and hence are not subject to economically motivated return migration that could bias the estimates.

A related consideration is that ethnic variation in the immigrant flow to the United States during this period was mainly driven by home country conditions (political instability, persecution), and hence the skill composition of that flow is unlikely to be the response to local economic conditions in the US. For example, the ethnic and skill mix of immigrants in 1920 were driven in large part by World War I, which made departure from combatant countries more difficult, particularly for individuals in low-wage occupations. Then, changes in the fraction of low-skilled immigrants arrived in 1915-1919 relative to that fraction in 1905-1909 constitute a largely exogenous source of variation in the difference between the average skills of adults in each ethnic group in 1920 relative to 1910. This is similar to the reasoning behind Angrist's (2002) study of the effects of sex ratios on marriage rates and labor market outcomes.[10]

---

unfeasible. For evidence on the suitability of the "linear in group means" model of peer effects in a more general setting, see Sacerdote and Marmaros (2005).

[10] The instrument in Angrist (2002) was constructed from the sex mix, not the occupational mix, of recorded immigrant flows by ethnicity and year.

Although omitted variables bias is the main motivation for my IV strategy, it is important to note that the OLS estimates of equation (4) may also be confounded by the fact that one regressor, $\bar{x}_{jt}$, is in fact close to being an average of another regressor, x_{ij}. In other words, the parents of children in my data are among the adults used to compute the average measure of skills in that ethnic group.

Suppose initially that $\bar{x}_{jt}$ was exactly the ethnic group mean of x_{ij}, then OLS estimates of the coefficient on $\bar{x}_{jt}$ in equation (4) would be equivalent to the augmented regression form of a Hausman (1978) specification test for the difference between OLS and IV estimates of the coefficient on x_{ij} in a simple regression of y_{ij} on x_{ij} only, with ethnicity dummies serving as the instrument for x_{ij}.[11] If OLS estimates differ from the IV estimates in the bivariate regression for any reason (e.g., measurement error in x_{ij}), then the estimated OLS coefficient on $\bar{x}_{jt}$ in equation (4) would be nonzero (positive, in the errors in variables case) even in the absence of ethnic peer effects. This problem is common to a broad class of empirical exercises where an outcome variable is affected by both an average and an individual level variable, and appears in the estimation of human capital externalities (Acemoglu and Angrist (2000)).

The situation here is somewhat more complicated since $\bar{x}_{jt}$ is not the exact average of x_{ij} in the sample: the average $\bar{x}_{jt}$ also includes the foreign born who do not have children, as well as all second generation adults of working age. Nevertheless, I show in Appendix A that under similar circumstances the estimate of β_2 will be non-zero even if the actual ethnic capital effect is zero. This problem is solved by treating both x_{ij} and $\bar{x}_{jt}$ as endogenous in (4).

To construct instruments for x_{ij}, I use father's age at arrival, interacted with a non-English speaking country of birth dummy, as an instrument for x_{ij}. Proficiency in the dominant language of the receiving country is a particularly important component of an immigrant's work-related human capital. Because languages are easier to learn at an earlier age, an immigrant who arrived as a child from a non-English speaking country should have developed better English-language skills than one who arrived as an adult. In

[11] See Davidson and MacKinnon (1989) for a detailed derivation of the equivalence between the Hausman specification test and its augmented (or 'artificial') regression form.

several studies of immigrants' language skills and earnings in Australia, Canada, Israel and the US, Chiswick and Miller (1992, 1995) and Miller and Chiswick (2002) report that, holding observable characteristics constant, language proficiency increases with years in the receiving country and is lower when immigrants have migrated at older ages. Research on cognitive science has established that the age of acquisition of a first or second language is a major determinant of ultimate proficiency (Newport (1990), Flege, Munro and MacKay (1995)).[12]

Since immigrants originating in English-speaking countries do not face a new language upon arrival to the US, these effectively serve as a control. With my strategy, only differences in outcomes between, say, two children of the same age whose respective fathers immigrated from Germany at different ages, net of differences in outcomes for comparable children whose fathers arrived from England at parallel ages, are attributed to parental capital. A similar strategy was used in Bleakley and Chin (2003) to study the returns to English proficiency for US immigrants.

C. Previous Research on Ethnic Peer Effects

Most empirical research on ethnic peer effects to date looks at the intergenerational transmission of socioeconomic outcomes such as schooling and earnings. One of the earliest empirical studies is Borjas' (1992) analysis of 1970s and 1980s General Social Surveys and National Longitudinal Surveys of Youth data. He regressed education and log wages of second-generation individuals on the education of their fathers, and on the average of parents in the ethnic group, and found that ethnic capital plays as important a role as the father's skills in determining the human capital of the next generation. In essence, he estimated an equation similar to (4) and interpreted the OLS coefficients as the causal effects of parental and ethnic capital. Borjas (1995) improves upon the previous study in addressing the potential problems introduced by measurement error in x_{ij}, by using sibling's reports of parental skills as instruments, but

[12] This is usually linked to the fact that puberty is associated with a biological reduction in the plasticity of the neural circuits that determine language learning ability (Lenneberg (1967), Flege, Yeni-Komshian and Liu (1999)).

still treats the level of skills in the ethnic group as exogenous and is therefore subject to omitted variable bias, as described in the previous section.[13]

A related set of papers seeks to estimate the intergenerational transmission parameter describing how the mean human capital of the ethnic group changes over time. When parental level of skills is not observed, it is possible to aggregate (5) and write:

$$y_{ijt} = \beta_0 + (\beta_1 + \beta_2)\,\bar{x}_{jt} + z_{it}'\gamma + \nu_{ijt} + \varepsilon_{ijt},$$

in which case it is only feasible to recover (β_1 + β_2), an 'intergenerational correlation coefficient,' but not the ethnic peer effect β_2 separately from the parental effect β_1. Borjas (1993), using 1940 and 1970 Census data; Borjas (1994), using 1910, 1940 and 1980 Census data; and Card, DiNardo and Estes (2000), using the 1940 and 1970 Censuses along with the 1994-96 CPS, all found intergenerational correlations of education and earnings in the range of 0.4 to 0.6. While this is an interesting question in and of itself, the exclusion of father's skills in the regressions makes it impossible to disentangle the ethnic peer effect from the direct transmission of skills within the family.

One additional caveat that applies to most of these studies is the fact that the measure of ethnic capital is generally constructed as the average skills in the parents' generation thirty years prior (when many of the individuals observed in the next generation had not even been born yet). The resulting estimates may be especially prone to bias from measurement error and economically motivated return migration. My strategy uses characteristics of the ethnic group actually faced by the children of immigrants when growing up.

III. Data and Main Results

A. Data Sources

The data used here comes from the 1910 and 1920 Census IPUMS files (documented in Ruggles and Sobek, 1997). Because information on the skills of parents is only available for the subsample of persons who still reside with their parents, which is

[13] To be precise, Borjas (1995) also experiments with using sibling's reported ethnicity as an instrument for ethnic capital. While that strategy corrects for measurement error in the assignment of the individual to an ethnic background, it does not address the potential omitted variables problem (from return migration or ethnicity-specific shocks, as explained above) that contaminates the measured level of ethnic capital in each group.

unlikely to be a representative subsample of adults, I restrict my analysis to an extract of second-generation children of schooling age (6 to 18 years old) and their parents. For the construction of measures of average skills by ethnic group, I use an extract of foreign-born and second-generation adults of working age (19 to 65 years old). The 1910 and 1920 Censuses contain detailed information on the age at arrival of immigrants, essential for the construction of one of my instruments, and on father's and mother's country of birth and mother tongue. The latter are used to classify ethnic groups in a manner similar to that employed in administrative data on immigrant arrivals by occupation and 'country or people' (Ferenczi and Willcox, (1929)) used to form the instruments. This results in twenty-six groups, plus an additional not elsewhere classified group. (See Appendix B for details on the coding scheme).

The outcome variable of interest is school attendance, the only education variable available for children but also perhaps the most relevant schooling measure in the early 20th century, before the 'high school movement' that made enrollment in secondary schooling more widespread (Goldin (1993)). Skills of parents and average skills in the ethnic group are measured using a proxy for literacy in English.[14] Literacy in English is represented by a dummy variable that equals one if an individual indicates he or she can read and write in some language *and* he or she can speak English. This way I aim to capture an informative dimension of human capital that is presumably valued in the labor market. As argued in the previous section, language proficiency is an important component of immigrants' skills as valued in the US labor market. The use of English literacy (henceforth referred to simply as literacy) to define ethnic capital facilitates the use of age at entry as an instrument for parental capital.

Table 2 gives descriptive statistics for the extract. While the distribution across regions and the average age of children and fathers remains fairly stable between 1910 and 1920, the fraction of second generation children attending school increases by about 3 percentage points, while the average literacy rate of immigrant fathers decreases slightly by almost 2 points. On the other hand, the average literacy of all first and second

[14] Even though a literacy requirement for immigration to the US existed since the Federal Act of March 3, 1893, it was hardly enforced and never became an effective system to screen out low-skilled aliens, thus keeping US immigration until 1921 virtually unrestricted: 23.5% of all immigrants admitted in 1900-1909, 22.1% of those admitted in 1920-14, and 10.0% of those admitted in 1915-19 were illiterate, according to official government records (US Census Bureau, 1975).

generation working age adults (my measure of ethnic capital) is higher in 1920 than in 1910, likely as a result of the higher average age (one and a half years older on average), the progressive accumulation of skills by previous immigrants and the higher proportion of second generation individuals over time. Finally, the high (relative to its time) fraction living in a metropolitan area reflects the fact that immigrants are disproportionately more likely to settle in urban areas than natives.

The existence of ethnic peer effects is the result of exposure to other individuals in the group who act as role models and have an influence on the skill acquisition of children. One way to measure the degree of interaction among individuals in a given community is by looking at endogamy rates. I use information on the nativity of spouses of married first and second generation women in order to compute the probability of marriage to an individual from the same (first or second generation) ethnic group, conditional on being married. The importance of intra-ethnic marriage in the groups defined in my sample is documented in Table 3, which reports the distribution of husbands' ethnicity separately for foreign born and second generation women. Endogamy rates are high for almost all groups even in the second generation, which suggests a strong level of individual interaction within groups. Over 80 percent of Italian women in the second generation married in the same group, and that percentage is even higher for Jewish and Japanese daughters of immigrants. In English-speaking groups such as English/Welsh or Irish, these rates are lower, yet only half of English, and only a third of Irish women of second generation have a native husband.[15] Table 3 therefore supports the ethnic taxonomy used in this analysis.

The ethnicity and skill distribution of the foreign stock (first and second generation individuals) are described in Table 4. There is a good deal of heterogeneity across ethnic groups and over time both in adult literacy and in children's school enrollment rates. This variability is more clearly reflected in Figure 1, which plots school attendance of second generation children against the average literacy rate of first and second generation adults for all 54 ethnicity-year cells. The figure shows that higher average literacy rates are associated with higher school enrollment rates for children.

[15] Moreover, breaking down the 'married other group' column would show that, in almost all cases, the endogamy rate for second generation women is still above the fraction of women who married in any other single group.

Next I will turn to regression analysis in order to control for individual characteristics as well as ethnicity and year effects, and then to use instrumental variables to identify what part of this observed relationship is caused by ethnic peer effects.

B. OLS Estimates

The estimating equation for second-generation individual *i*, in ethnic group *j*, observed in Census year *t* is (4), derived in the previous section. The first stage equations relate the endogenous regressors to the instruments a_{ijt}, father's age at arrival interacted with a dummy for non-English speaking country of birth, and f_{jt}, the fraction of 'low-skilled' immigrants (laborers, servants and agricultural workers) arrived in the five years prior to the Census year:

$$x_{ijt} = \alpha_1 + \rho_{11} a_{ijt} + \rho_{12} f_{jt} + \delta_{1j} + \delta_{1t} + z_i' \gamma_1 + \varepsilon_{1ijt} \text{ , and} \quad (5)$$

$$\bar{x}_{jt} = \alpha_2 + \rho_{21} a_{ijt} + \rho_{22} f_{jt} + \delta_{2j} + \delta_{2t} + z_i' \gamma_2 + \varepsilon_{2ijt} \text{ .} \quad (6)$$

This system is just-identified. The covariates z_i include region effects, age, father's age, and other demographics. Note that I also include a father's age-at-arrival main effect in the equation of interest. Even though the immigration decision of the father is previous to the birth of the child in my sample, and therefore could be thought of as exogenous to children's outcomes, the timing of the father's arrival to the US may be correlated with unobserved parental characteristics such as ambition and drive, which may then be transmitted to the next generation. I allow father's age at arrival to enter the equation and directly affect schooling of children.

Table 5 reports OLS estimates of equation (4). These suggest that parents' literacy has a modest but precisely measured effect on school attendance of children, while the average literacy in the ethnic group has a relatively large and significant impact.[16] While region of residence and metropolitan area do not appear to affect the estimates for parental capital, the ethnic peer effect declines notably (from 0.215 in column 1 to 0.135 in column 4) after including other controls such as age, father's age, number of siblings, and father's age at arrival. In order to address the potential shortcomings According to these results, two comparable children who only differ in the literacy of their fathers are

[16] Standard errors in all regressions are corrected for ethnicity-year clustering.

predicted to have a difference in the probability of attending school of about 5 percentage points. Two observationally equivalent children with equally skilled parents but belonging to ethnic groups that differ in their literacy rates by 30 percentage points are predicted to differ in their respective probabilities of school attendance by just over 4 percentage points.

Because the linear model does not impose that predicted probabilities fall within the zero-one interval for all observations, it might not be an accurate representation of the relationship between father's literacy, ethnic capital and school attendance. For that purpose, I consider an alternative, nonlinear specification of equation (4), namely a probit model in which the right-side variables are the determinants of a latent index governing the school attendance decision. Estimates of the marginal effects implied by the probit model, which are reported in column 5, are very similar to the OLS results, indicating that the linear probability model is a useful approximation. Column 6 experiments with using an average of father's and mother's literacy, to account for the role of both parents in the transmission of skills.[17] The results are comparable to those in the previous columns: even though ethnic spillovers are estimated to be slightly lower, they still amount to twice the parental effect.

The estimates in this table are not readily comparable to other estimates in the literature. They are most relatable to Sacerdote (2005), who in his analysis of the transmission of human capital between former slaves and their children and grandchildren reports that having a mother who was born a slave decreases the probability of being enrolled in school by 12 percent, and to Weir (2000), who reports positive effects of parents' years of schooling on school enrollment of children. I am not aware, however, of any studies of the intergenerational correlation between immigrant parents and second-generation children that have looked at literacy as a regressor and school enrollment as an outcome variable.

As noted in the previous section, OLS estimates of ethnic capital effects are subject to upward bias from measurement error in father's skills. In that case, not only does the measurement error attenuate the coefficient on parental capital, but it can also

[17] The literature usually finds similar results when child outcomes are correlated with mother's characteristics (see Card, DiNardo and Estes (2000) or, for a cross-ethnic study, Gang (1996)).

create a false impression of positive ethnic peer effects. To illustrate the implications of an inconsistent estimate of parental effects for the identification of ethnic effects, I estimated equation (4) imposing different plausible values for β_1. As reported in columns 2 and 3 of Table 6, the estimated peer effect is 0.144 when father's literacy is excluded from the equation, but falls to 0.106 when the parental effect is set to 0.20. On the other hand, changing the constrained value of the ethnic spillover does not have much impact on the estimated parental effect. These results support the notion that measurement error in parental skills can bias the estimation of ethnic peer effects, and therefore it is fundamental to estimate β_1 consistently in order to identify β_2.

The first-stage estimates for father's literacy rate (from estimating equation 6) are displayed in Table 7. There is a strong, negative relationship between the instrument a_{ijt} and parental skills. Regardless of the controls used, the estimate implies that delaying arrival from a non-English speaking country to the US by three years leads to a two percentage point decline in the probability of speaking English and being literate.[18] Table 8 reports a set of first-stage estimates for average literacy. Even though the instrument is later used in a micro regression on tens of thousands of observations, it is insightful to estimate equation (6) at the aggregate level, controlling for ethnicity and year main effects only, given that both the endogenous regressor ($\bar{x}_{jt}$) and the instrument (f_{jt}) do not vary within ethnicity-year cells. Column 1 shows that the fraction of low-skilled recent arrivals does have a large, significantly negative effect on average literacy rates even at the macro level, on only 54 observations corresponding to the ethnicity-year cells. The point estimate reveals that a 10 percentage point rise in the fraction of new immigrants with low skills in a given ethnic group leads to a 6 percentage point decline in average literacy rates in that group. This negative relationship is illustrated in Figure 2, which plots literacy rates and fractions of low-skilled recent immigrants, net of ethnicity and year. Columns 2 through 5 confirm that the estimate is robust to the inclusion of controls at the micro level. Interestingly, neither a_{ijt} comes in significantly in equation (1-6) nor

[18] These results are not directly comparable to those in Bleakley and Chin (2003). The English proficiency variable in the 1990 Census, which they use in their estimations, is coded into four different categories, whereas my measure of skills is a binary variable. It is also worth pointing out that I experimented with the non-linear function of age at arrival that Bleakley and Chin use in their definition of the instrument a_{ijt}, and obtained very similar results. The non-linearity likely becomes important only in distinguishing between subtle differences in language proficiency, but does not matter in predicting my binary skill indicator.

does f_{jt} in equation (6), confirming that each instrument is a strong predictor only of one endogenous variable, along the lines of the discussion on the identification strategy outlined above.

C. IV Estimates

The two-stage least squares (2SLS) estimates of equation (4) are reported in Table 9. The coefficient on father's literacy after adding all the controls, as shown in column 3, is clearly higher than its OLS counterpart. The results indicate that, other things equal, having a literate father increases the probability that a child is enrolled in school by 20 percentage points. The OLS estimate from Table 5 appears to be downward biased, which is consistent with measurement error in the measure of parental skills. At the same time, models that treat the average skills in the ethnic group as endogenous generate a 2SLS estimate of 0.116 for the effect of ethnic capital on the probability that children are in school. Moreover, experimenting with a probit specification where both parental and ethnic capital are treated as endogenous produces highly comparable estimates, as shown in column 4. Therefore, it appears that most of the positive association observed in Table 5 and Figure 2 is indeed causal. The point estimates for ethnic spillovers are, however, slightly lower than the OLS ones, which is coherent with the omitted variables bias story whereby some ethnic groups experience positive shocks that encourage further skill accumulation and result in both higher literacy levels of adults and higher school enrollment rates of children. This difference between the OLS and 2SLS estimates of ethnic peer effects is, nevertheless, not significant. Using mean skills of both parents yields slightly lower but less precise estimates (due mostly to reduced sample size). In any case, the pattern of the estimates relative to their OLS counterparts is in line with that of all other columns, reinforcing the idea that measurement error in parental skills is a severe problem.

Table 10 performs the same experiment as in Table 6, now using instrumental variables to estimate the unconstrained coefficient. As before, the coefficient on ethnic capital shrinks when the parental effect is larger. When the latter is set to 0.20, approximately the 2SLS result from the previous table, the estimated ethnic spillover becomes equal to the unconstrained 2SLS result. This provides further proof that

consistent estimation of one endogenous regressor is key to the correct identification of the other.

D. Additional Results and Specification Checks

I turn now to addressing the concern that my results are affected by the multi-dimensional nature of human capital. Suppose that the skill that is transmitted from one generation to the next is not a single factor, but instead comprises two different components, x_1 and x_2: $x_{ijt} = x_{1ijt} + x_{2ijt}$. Only one component, x_1 (literacy), is observable. In that case, the estimating equation (4) becomes:

$$y_{ijt} = \alpha + \beta_1 x_{1ijt} + \beta_2 \bar{x}_{1jt} + \delta_j + \delta_t + z_i'\gamma + \left(\varepsilon_{ijt} + \beta_1 x_{2ijt} + \beta_2 \bar{x}_{2jt}\right). \qquad (7)$$

Since only x_1 can be included in the regression, the residual contains x_2 and $\bar{x}_2$ [19]. Given that f_{jt}, my instrument for (observed) ethnic capital is based on the occupational mix of new immigrants, it may be picking up some unobservable dimensions of skills that are not included in literacy. In that case, f is correlated with $\bar{x}_2$ and hence with the residual, thus yielding inconsistent estimates. Including father's occupation in the regression, however, should control for that additional component of transferable skills not contained in the observed x_1. The first two columns in Table 11 report OLS and 2SLS models that include a dummy variable that equals one if the father is a laborer, a servant or an agricultural worker (the same criterion used to construct the instrument *f*). If unobserved skills correlated with f rendered my instrumental variable strategy invalid, then these estimates should be different from those that do not include a proxy for unobserved skills. There is, however, no evidence that the inclusion of father's occupation alters the estimates in any way. These results therefore strengthen the case for interpreting the 2SLS estimate as the causal effect of ethnic capital, as measured by average literacy, on school attendance of children.

Another possible concern is the fact that, by focusing on individuals between 6 and 18 years of age, the sample may be pooling young children whose schooling is solely determined by their parents with teenagers whose school attendance behavior is relatively

[19] To the extent that different components of skills may be correlated, that alone creates an additional source of bias for OLS estimates of equation (7).

more likely to reflect their own decisions. Excluding 17- and 18-year-olds from the estimation, however, does not lead to different results.

Finally, columns 5 and 6 in Table 11 deal with the possibility that imprecise estimation of averages for small ethnic groups may be biasing my estimates. For that purpose, I re-estimate equation (4) after excluding observations belonging to the following groups: African, Spanish, Romanian, Armenian and Ruthenian (the five smallest ethnic groups in my sample, as evidenced by the counts in Table 3). Again, my findings also survive this robustness check.[20]

IV. Ethnic Capital Effects and Measures of Ethnic Concentration and Interaction

The ethnic peer effects hypothesis has a number of implications that can be checked. First, ethnicity is likely to play a more important role among individuals who grow up in an environment with a higher concentration of people in their ethnic group. In regions where one's ethnic group only represents a very small fraction of the population, children will probably be exposed to, and influenced by, less frequent social and cultural intragroup contacts. The analysis in the preceding section ignored this because it assumed that the coefficient on ethnic capital was the same across individuals. In order to explore whether ethnic clustering affects the magnitude of ethnic spillovers, I interact average skills with a measure of concentration in the ethnic group. For each child in my sample, I compute the proportion of working age adults in the region who share the same ethnic background. A dummy variable indicating whether that fraction is above or below the average across observations is interacted with both the parental capital variable and the ethnic capital variable.[21] Admittedly, a sharper exercise would compare individuals in highly segregated neighborhoods against those in more homogeneous districts. The 1910

[20] I also experimented with excluding Germans, given possible concerns about their increased unpopularity during this period, as well as with the existence of schools where German was the language of instruction. The point estimates remained largely unchanged.

[21] The average fraction of the working age adult population in the same ethnic group as the child in my sample is just under 12%. I therefore define my dummy variable for 'high' ('low') concentration as being in a region with more (less) than 12% of adults in the same ethnic group. In order to compute that fraction, I look at both first and second generation adults aged 19 to 65 (which are the most likely to interact with the parents of the child).

and 1920 Census data, however, does not include such detail of information on place of residence, so I use Census region instead.[22]

The findings are summarized in Table 12. Column 1 reproduces the baseline OLS estimates from column 4 in Table 5. Column 2 shows that the ethnic peer effect is larger among persons who live in highly concentrated areas (0.261 versus 0.122 for children in low concentration regions), even though the standard errors are too high to claim the difference is statistically significant. The loss in precision occurs because not all ethnic groups are represented in both high and low concentration regions, and hence estimation of each of the parental and capital effects no longer uses all ethnicity-year cells. The last two columns repeat the same exercise for 2SLS. While column 3 shows the benchmark 2SLS estimates from column 3 in Table 9, Column 4 reports the coefficients separately for high and low concentration areas. Again, despite the loss in precision, the coefficient on ethnic capital is higher where concentration is greater. These results are suggestive that ethnic environment has a stronger impact on children in areas where ethnic groups are more concentrated.

Another check on the peer effects story looks at differences in the magnitude of the coefficient on ethnic capital as a function of a different measure of social interactions within groups. As has been argued in Section II, endogamy rates provide a good measure of the extent to which individuals in an ethnic group are in close contact to other people in the group as opposed to people in other groups. Communities where most women marry within their ethnic group are typically more cohesive and closed to outside influences. On the other hand, children in those communities where a large proportion of women marry outside their ethnic group are more likely to interact with neighbors or relatives of different ethnicities, and should be less frequently exposed to the particular role models and values associated with their own ethnic group. If that hypothesis is correct, ethnic peer effects in more endogamous communities must be stronger than in less endogamous groups.

To determine whether ethnic peer effects are associated with high endogamy rates, I allow the coefficient on ethnic capital to vary according to the fraction of married

[22] I do not use state of residence, because the number of first and second generation adults of working age by state in 1910 is too small and introduces too much sampling error in the measures of concentration.

second-generation women in the region who wedded in the same ethnic group. I use second generation women because endogamy rates for the first generation might simply reflect the fact that many immigrants married before arrival to the US, whereas the marriage decisions of their US-born children provide a more accurate measure of the actual level of interaction among members of the same ethnic group.[23]

Table 13 reports regressions where father's literacy and average literacy in the ethnic group are interacted with dummy variables indicating whether the endogamy rate in the region was above or below 55 percent, which is roughly the average second-generation endogamy rate in the sample. OLS estimates in column 2 indicate that ethnic capital externalities are larger in highly endogamous ethnic groups. This is further confirmed by the 2SLS estimates in column 4. The estimated ethnic peer effect is insignificant and very close to zero for those in low endogamy communities, and 0.140 for those in high endogamy groups.

Aside from providing further support to my ethnic group classification, these results imply that ethnic spillovers operate mainly through the strength of the ethnic social fabric, as measured by endogamy rates. There is evidence that as cultural and socioeconomic assimilation takes place, cross-ethnicity marriage rates increase and endogamy rates decline (Spickard (1989)). Those communities with both few endogamous unions and low ethnic spillovers are thus likely to be more integrated in the US. In such groups, then, exposure to ethnic role models and behavioral norms becomes more infrequent, and the importance of ethnic capital in the transmission of skills across generations diminishes.

V. Conclusions

Previous attempts to identify the link between average skills of immigrants and the socioeconomic outcomes of their children have paid little attention to problems of omitted variables bias and measurement error. My research underscores the potential importance of endogenous ethnic and parental skills in intergenerational skill transmission equations and of their sensitivity to the estimation procedure used in the analysis.

[23] Endogamy rates for second-generation women were presented in column 5 of Table 3.

Estimates using an exogenous source of variation in skills among immigrant groups, while simultaneously instrumenting for the skills of parents to reduce attenuation bias, provide strong evidence for the existence of ethnic capital effects, albeit not of a stronger magnitude than the direct effect that parents have on their children. Additional specification checks support the notion that ethnic peer effects operate partly through the geographic concentration of ethnic groups and the higher level of interaction among individuals in those groups.

The persistence of ethnic differentials across generations and over time has relevant implications for welfare and immigration policy. While the outcome variable studied here is school enrollment, the estimated ethnic capital effects have far-reaching consequences. A lower probability that a child attends school implies reduced opportunities for social mobility and ultimately translates into lower earnings. The existence of ethnic peer effects in the human capital accumulation process of children has long-lasting effects on inequality, and shows that incumbent ethnic communities are correct to be concerned about the dilution of skills resulting from the arrival of new immigrants to the group. On the other hand, it also indicates that government interventions in the form of aid programs specifically targeted at particular ethnic groups can be a very effective means to reduce inequality in the short *and* in the long run, for that same reason.

Appendix

A. Mathematical Appendix

This section attempts to develop more formally the point that a positive estimated coefficient on ethnic capital can be obtained even in the absence of ethnic peer effects. Consider a simplified version of equation (4), where $\beta_2 = 0$ and all covariates have been dropped or 'partialed out.' Moreover, assume that x is only a noisy measure of the true parental skill variable, x^*. The model then becomes:

$$y_i = \beta_1 x_i^* + \varepsilon_i, \tag{A1}$$

where $x_i = x_i^* + \nu_i$, the measurement error term ν_i has mean zero and variance σ_ν^2, and it satisfies $E[\nu_i x_i^*] = E[\nu_i \varepsilon_i] = 0$ (classical measurement error).

In addition to x, another variable w is available such that $E[w_i x_i^*] \equiv \sigma_{wx} > 0$ and $E[w_i \varepsilon_i] = 0$. Without loss of generality, then, this new variable w is positively correlated with parental skills. A regression that includes both x and w will yield:

$$y_i = \pi_1 x_i + \pi_2 w_i + u_i, \tag{A2}$$

with:

$$p\lim \pi_2 = \beta_1 \left[\frac{\sigma_{wx}(1-\lambda)}{\sigma_w^2 - \lambda(\sigma_{xw}^2/\sigma_x^2)} \right], \tag{A3}$$

where $\sigma_w^2 = E[w_{ij}^2]$ and $\sigma_x^2 = E[x_{ij}^2]$ (all variables are measured in deviations from their means) and $\lambda = [\sigma_x^2/(\sigma_x^2 + \sigma_\nu^2)]$, or the 'reliability ratio', a measure of the goodness of x as a measure of x^*. Since $0 < \lambda < 1$, the coefficient on w in this regression does not converge to zero. In words, the introduction of an additional regressor that is correlated with the mismeasured parental capital results in biased coefficients and the misleading appearance that the new regressor 'belongs' in the equation, when in fact it is not present in the true model (A1). The sign of the probability limit of π_2 is that of the covariance between w and x (positive). Of course, if no measurement error is present, then π_2 is asymptotically zero.

Ethnic capital is an example of such a regressor w. To be more precise, consider $w_i = x_i^* + \eta_i$, where η_i has mean zero and variance σ_η^2, and $E[\eta_i x_i^*] \equiv \sigma_{x\eta}$ which does

not necessarily equal zero. In that case, $\sigma_{wx} = \sigma_x^2 + \sigma_{x\eta}$ and $\sigma_w^2 = \sigma_x^2 + \sigma_\eta^2 + 2\sigma_{x\eta}$, so (A3) becomes:

$$p\lim \pi_2 = \beta_1 \left[\frac{\left(\sigma_x^2 + \sigma_{x\eta}\right)(1-\lambda)}{\left(\sigma_x^2 + \sigma_{x\eta}\right)\left[1 - \lambda\left(\frac{\sigma_x^2 + \sigma_{x\eta}}{\sigma_x^2}\right)\right] + \sigma_\eta^2 + \sigma_{x\eta}} \right]. \qquad \text{(A4)}$$

When *w* does not vary within a group *j*, but only across groups (this is, when $w_{ij} \equiv \bar{x}_j = x_{ij}^* + \eta_{ij}$), then $\sigma_{x\eta} = -\sigma_\eta^2$ (intuitively: the covariance must be negative because relatively high realizations of x^* will require relatively low values of η in order for all observations in group *j* to have the same value of *w*). Hence $\sigma_w^2 = \sigma_x^2 - \sigma_\eta^2$ and therefore the coefficient will converge to:

$$p\lim \pi_2 = \beta_1 \left[\frac{1-\lambda}{1 - \lambda\left(\frac{\sigma_w^2}{\sigma_x^2}\right)} \right]. \qquad \text{(A5)}$$

Since $\sigma_w^2 < \sigma_x^2$, the above formula is bounded between 0 and β_1.

The derivations above show that the OLS estimate of the coefficient on ethnic capital ($\bar{x}_j$) in equation (4) is inconsistent, because $\bar{x}_j$ is some ethnicity-specific summary measure of skills that is correlated with x_{ij}, even if it is not the exact average of the fathers in the sample. One should then expect the coefficient to be positive even in the absence of ethnic spillovers, just from the fact that the ethnic mean is correlated with the true parental skills, which are observed with error.

Finally, note that in the particular case where $\bar{x}_j$ is actually computed for each ethnic group as the average of x_{ij} in the sample, then the term $\left(\sigma_w^2/\sigma_x^2\right)$, or $\left(\sigma_{\bar{x}}^2/\sigma_x^2\right)$, in (A5) can be read as the R-squared of the first stage regression of *x* on a full set of ethnicity dummies. The better the fit in that first stage (this is, the better ethnicity predicts x_{ij}), the larger the bias, and the stronger the spurious 'ethnic capital effect' will appear to be.

B. Data Appendix

Ferenczi and Willcox (1929) report administrative data on alien arrivals collected by the United States immigration authorities. This source shows numbers of immigrants admitted annually by broad occupation categories (agriculture, laborers and servants, professionals, commerce and finance, industry, and miscellaneous) and by "race or people," for every year between 1899 and 1924. Additional tables classify immigrants by "race or people" into their countries of origin, which is the information I used to match the groups in the administrative data to the ethnic groups I identified in the Census data. Every Census from 1880 to 1970 collected information on country of birth that identifies the foreign born, and the foreign-birth status of both parents. Moreover, the nativity variables were recoded in the IPUMS to give a fairly consistent categorization for all years.

Classification of the foreign born (first generation) individuals in my sample into ethnic groups was made using country of birth or a combination of country of birth and mother tongue or race, in order to match the "race or people" categories in Ferenczi and Willcox as closely as possible. The coding scheme was as follows: [When different, the Ferenczi and Willcox categories appear in brackets.]

1. African (Black): born in Africa or the West Indies, and of black race.
2. Armenian: born in the former Russian Empire/Soviet Union or the Middle East, with Armenian as their mother tongue.
3. Bulgarian/Serbian/Croatian/Slovenian: born in Bulgaria or the former Yugoslavia, or elsewhere in Central/Eastern Europe with Bulgarian, Serbo-Croatian or Slovene as their mother tongue. ["Bulgarian, Serbian and Montenegrin," and "Croatian and Slovenian," in Ferenczi and Willcox (1929)].
4. Czech: born in Bohemia or Moravia, or elsewhere in Central/Eastern Europe with Czech (Bohemian or Moravian) as their mother tongue. ["Bohemian and Moravian," in Ferenczi and Willcox (1929)].
5. Dutch/Flemish: born in the Netherlands, or in Belgium with Dutch (Flemish) as their mother tongue.

6. English/Welsh: born in England, Wales or British Canada, or in Canada with English as their mother tongue.[24]

7. Finnish: born in Finland.

8. French/Canadian: born in French Canada (Québec) or France, or in Canada,[25] Belgium or Switzerland with French as their mother tongue.

9. German/Austrian: born in German or Austria, or elsewhere in Central/Eastern Europe, with German as their mother tongue ["German," in Ferenczi and Willcox (1929)].

10. Greek: born in Greece.

11. Hungarian: born in Central/Eastern Europe, with Hungarian/Magyar as their mother tongue. ["Magyar," in Ferenczi and Willcox (1929)].

12. Irish: born in Ireland, or in British Canada with Irish as their mother tongue.

13. Italian: born in Italy.

14. Japanese: born in Japan.

15. Jewish: born in Central/Eastern Europe, with Yiddish as their mother tongue. ["Hebrew," in Ferenczi and Willcox (1929)].

16. Lithuanian: born in Lithuania or elsewhere in Central/Eastern Europe with Lithuanian as their mother tongue.

17. Mexican: born in Mexico.

18. Polish: born in Poland or elsewhere in Central/Eastern Europe, with Polish as their mother tongue.

19. Portuguese: born in Portugal, or in South America with Portuguese as their mother tongue.

20. Romanian: born in Romania, or elsewhere in Central/Eastern Europe with Romanian as their mother tongue.

21. Russian: born in the Russian Empire/Soviet Union, or elsewhere in Central/Eastern Europe, with Russian as their mother tongue.

22. Ruthenian: born in Central/Eastern Europe with Ruthenian as their mother tongue.[26]

---

[24] In 1910 all individuals born in Canada are given the same code for birthplace, hence the use of mother tongue to distinguish English from French Canadians.

[25] See previous footnote.

[26] Ruthenia (originally a Latin rendering of the ancient place name Rus) is a name applied to parts of Eastern Europe which were populated by Eastern Slavic peoples, as well as to various states that existed in this territory in the past. In the early twentieth century, the name 'Ruthenian' referred to an area then part of

23. Scandinavian: born in Norway, Iceland, Denmark or Sweden.

24. Scottish: born in Scotland, or in British Canada with Scottish Gaelic as their mother tongue.

25. Slovak: born in Slovakia, or elsewhere in Central/Eastern Europe with Slovak as their mother tongue.

26. Spanish: born in Spain, or in South America with Spanish as their mother tongue.

27. NEC: Not Elsewhere Classified.

Ethnicity of the second generation was assigned as above, but using father's country of birth and father's mother tongue, except for those with a foreign mother only, in which case mother's country of birth and mother's mother tongue were used.[27]

---

the Austro-Hungarian Empire which subsequently became the fourth province of interwar Czechoslovakia. Today the historical territory of Ruthenia comprises territories of present-day Belarus and Russia, the westernmost region of Ukraine, a small part of north-eastern Slovakia and a narrow strip of south-eastern Poland.

[27] I experimented with a definition of second generation based on the mother's country of birth and mother tongue, and using the father's information for those with a foreign father only. The results were not sensitive to the definition of the second generation.

References

Acemoglu, Daron, and Joshua D. Angrist (2000), "How Large Are Human Capital Externalities? Evidence from Compulsory Schooling Laws," *NBER Macroannual 2000*: 9-59.

Alba, Richard D. (1992), *Ethnic Identity: The Transformation of White America* (New Haven: Yale University Press).

Angrist, Joshua D. (2002), "How Do Sex Ratios Affect Marriage and Labor Markets? Evidence from America's Second Generation," *Quarterly Journal of Economics* 117(3): 997-1038.

Angrist, Joshua D., and Kevin Lang (2004), "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program," *American Economic Review* 94(5): 1613-1634.

Benabou, Roland J.M. (1993), "Workings of a City: Location, Education, and Production," *Quarterly Journal of Economics* 108(3): 619-652.

Bertrand, Marianne; Erzo F.P. Luttmer, and Sendhil Mullainathan (2000), "Network Effects and Welfare Cultures," *Quarterly Journal of Economics* 115(3): 1019-1055.

Bleakley, Hoyt, and Aimee Chin (2004), "Language Skills and Earnings: Evidence from Childhood Immigrants," *Review of Economics and Statistics*, 86(2): 481-496.

Borjas, George J. (1992), "Ethnic Capital and Intergenerational Mobility," *Quarterly Journal of Economics* 107(1): 123-150.

Borjas, George J. (1993), "The Intergenerational Mobility of Immigrants," *Journal of Labor Economics* 11(1): 113-135.

Borjas, George J. (1994), "Long-Run Convergence of Ethnic Skill Differentials: The Children and Grandchildren of the Great Migration," *Industrial and Labor Relations Review* 47(4): 553-573.

Borjas, George J. (1995), "Ethnicity, Neighborhoods, and Human Capital Externalities," *American Economic Review* 85(3): 365-390.

Card, David; John DiNardo, and Eugena Estes (2000), "The More Things Change: Immigrants and the Children of Immigrants in the 1940s, the 1970s and the 1990s," in George J. Borjas, ed., *Issues in the Economics of Immigration* (Chicago: University of Chicago Press).

Case, Anne C., and Lawrence F. Katz (1991), "The Company You Keep: The Effect of Family and Neighborhood on Disadvantaged Youths," NBER Working Paper No. 3705, Cambridge, MA.

Chiswick, Barry R., and Paul W. Miller (1992), "Language in the Immigrant Labor Market," in Barry R. Chiswick, ed., *Immigration, Language and Ethnicity: Canada and the United States* (Washington, DC: AEI Press).

Chiswick, Barry R., and Paul W. Miller (1995), "The Endogeneity between Language and Earnings: An International Analysis," *Journal of Labor Economics* 13(2): 246-288.

Coleman, James S. (1988), "Social Capital in the Creation of Human Capital," *American Journal of Sociology* 94(Supplement): S95-S120.

Davidson, Russell, and James G. MacKinnon (1989), "Testing for Consistency Using Artificial Regressions," *Econometric Theory* 5: 363-364.

Evans, William N.; Wallace E. Oates, and Robert M. Schwab (1992), "Measuring Peer Group Effects: A Study of Teenage Behavior," *Journal of Political Economy*, 100: 966-991.

Ferenczi, Imre, and Walter F. Willcox (1929), *International Migrations, Volume I: Statistics, Demographic Monograph No. 7* (New York: National Bureau of Economic Research).

Flege, James E.; Murray J. Munro, and Ian R. A. MacKay (1995), "Factors Affecting Strength of Perceived Foreign Accent in a Second Language," *Journal of the Acoustical Soceity of America* 97(5): 3125-3134.

Flege, James E.; Grace H. Yeni-Komshian, and Serena Liu (1999), "Age Constraints on Second-Language Acquisition," *Journal of Memory and Language* 41: 78–104.

Gang, Ira (1996), "Who Matters Most? The Effect of Parent's Schooling on Children's Schooling," Rutgers University, Department of Economics Working Paper No. 199613.

Glaeser, Edward L.; Bruce I. Sacerdote, and José Scheinkman (1996), "Crime and Social Interactions," *Quarterly Journal of Economics*, 91: 507-548.

Glazer, Nathan, and Daniel P. Moynihan (1963), *Beyond the Melting Pot: The Negroes, Puerto Ricans, Jews, Italians, and Irish of New York City* (Cambridge: MIT Press).

Goldin, Claudia (1998), "America's Graduation from High School: The Evolution and Spread of Secondary Schooling in the Twentieth Century," *Journal of Economic History* 58: 345-74.

Hausman, Jerry (1978), "Specification Tests in Econometrics," *Econometrica* 46: 1251-1271.

Howe, Irving (1976), *World of Our Fathers* (New York: Harcourt Brace Jovanovich).

Hoxby, Caroline (2000), "Peer Effects in the Classroom: Learning from Gender and Race Variation," NBER Working Paper No. 7867, Cambridge, MA.

Lenneberg, Eric H. (1967), *Biological Foundations of Language* (New York: Wiley).

Loury, Glenn C. (1977), "A Dynamic Theory of Racial Income Differences," in Phyllis A. Wallace and Annette A. LaMond, ed., *Women, Minorities and Employment Discrimination* (Lexington, MA: Lexington Books, 1977).

Lundberg, Shelly J., and Richard Startz (1998), "On the Persistence of Racial Inequality," *Journal of Labor Economics* 16(2): 292-323.

Manski, Charles F. (1993), "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies* 60: 531-542.

Miller, Paul W., and Barry R. Chiswick (2002), "Immigrant Earnings: Language Skills, Linguistic Concentrations and the Business Cycle," *Journal of Population Economics* 15(1): 31-57.

Newport, Elissa L. (1990), "Maturational Constraints on Language Learning," *Cognitive Science* 14(1): 11-28.

Putnam, Robert D. (1995), "Bowling Alone: America's Declining Social Capital," *Journal of Democracy* 6(1): 65-78.

Putnam, Robert D. (2000), *Bowling Alone* (New York: Simon & Schuster).

Ruggles, Steven, and Matthew Sobek (1997), *Integrated Public Use Microdata Series, Version 2.0, Volume 1: User's Guide*, (Minneapolis: Historical Census Projects, University of Minnesota Department of History).

Sacerdote, Bruce (2001), "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *Quarterly Journal of Economics* 116(2): 681-704.

Sacerdote, Bruce (2005), "Slavery and the Intergenerational Transmission of Human Capital," *Review of Economics and Statistics* 87(2): 217-234.

Sacerdote, Bruce and David Marmaros (2005), "How Do Friendships Form?," NBER Working Paper No. 11530, Cambridge, MA.

Sowell, Thomas (1981), *Ethnic America* (New York: Basic Books).

Spickard, Paul R. (1989), *Mixed Blood: Intermarriage and Ethnic Identity in Twentieth-Century America* (Madison: University of Wisconsin Press).

U.S. Census Bureau (1975), *Historical Statistics of the United States: Colonial Times to 1970* (Washington, DC: US Government Printing Office).

Weir, Sharada (2000), "Intergenerational Transfers of Human Capital: Evidence on Two Types of Externalities," Working Paper 2000.15, Centre for the Study of African Economies, University of Oxford.

Winston, Gordon C., and David J. Zimmerman (2004), "Peer Effects in Higher Education," in Caroline Hoxby, ed., *College Decisions: How Students Actually Make Them and How They Could* (Chicago: University of Chicago Press).

Zimmerman, David J. (2003), "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment," *Review of Economics and Statistics* 85(1): 9-23.

Figure 1
Children School Attendance versus Average Literacy in Ethnic Group.
No Ethnicity or Year Effects.

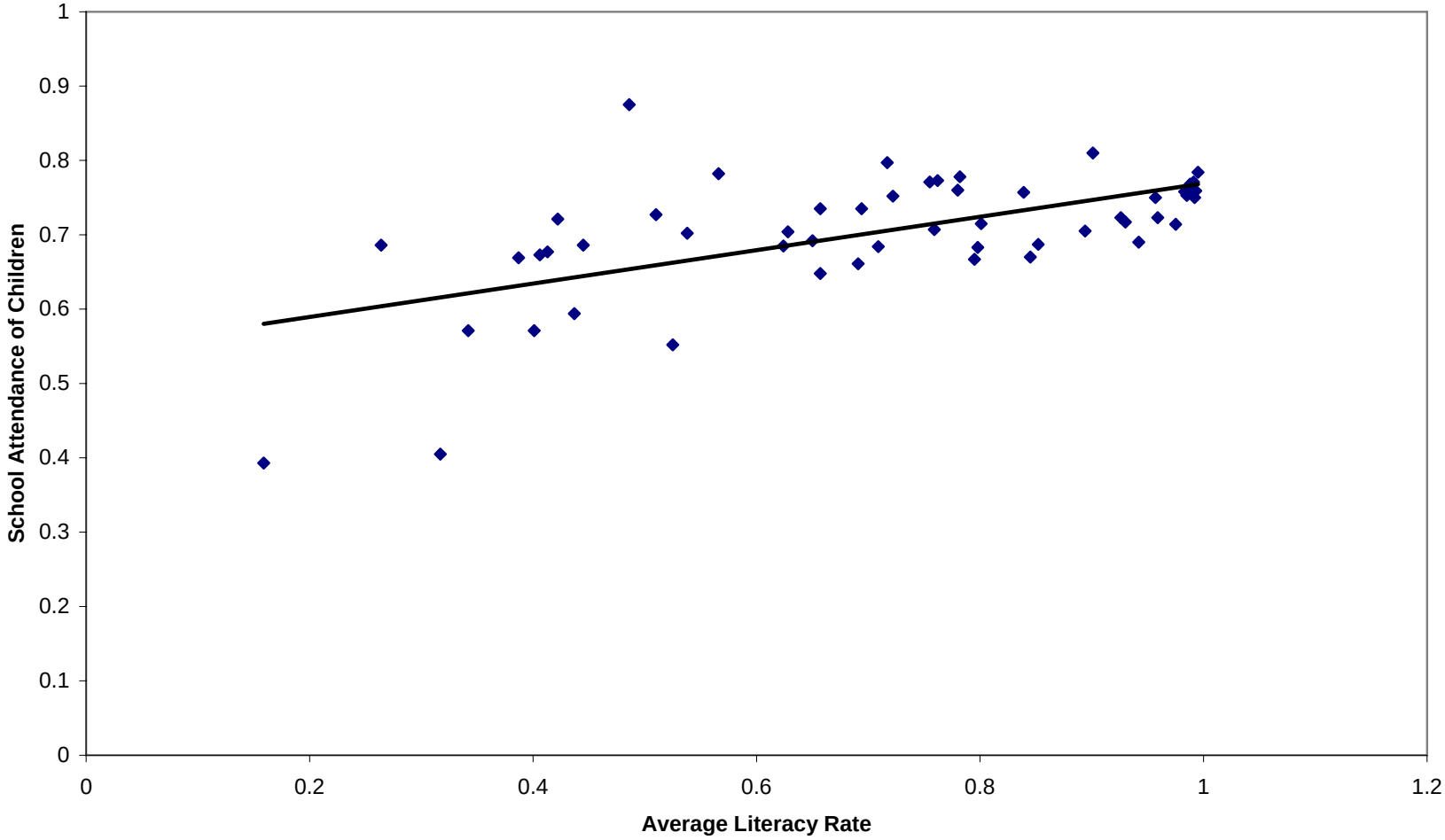

Figure 2
First Stage for Ethnic Capital: Average Literacy versus Fraction of Low-Skilled Recent Immigrants in Ethnic Group, with Ethnicity and Year Effects

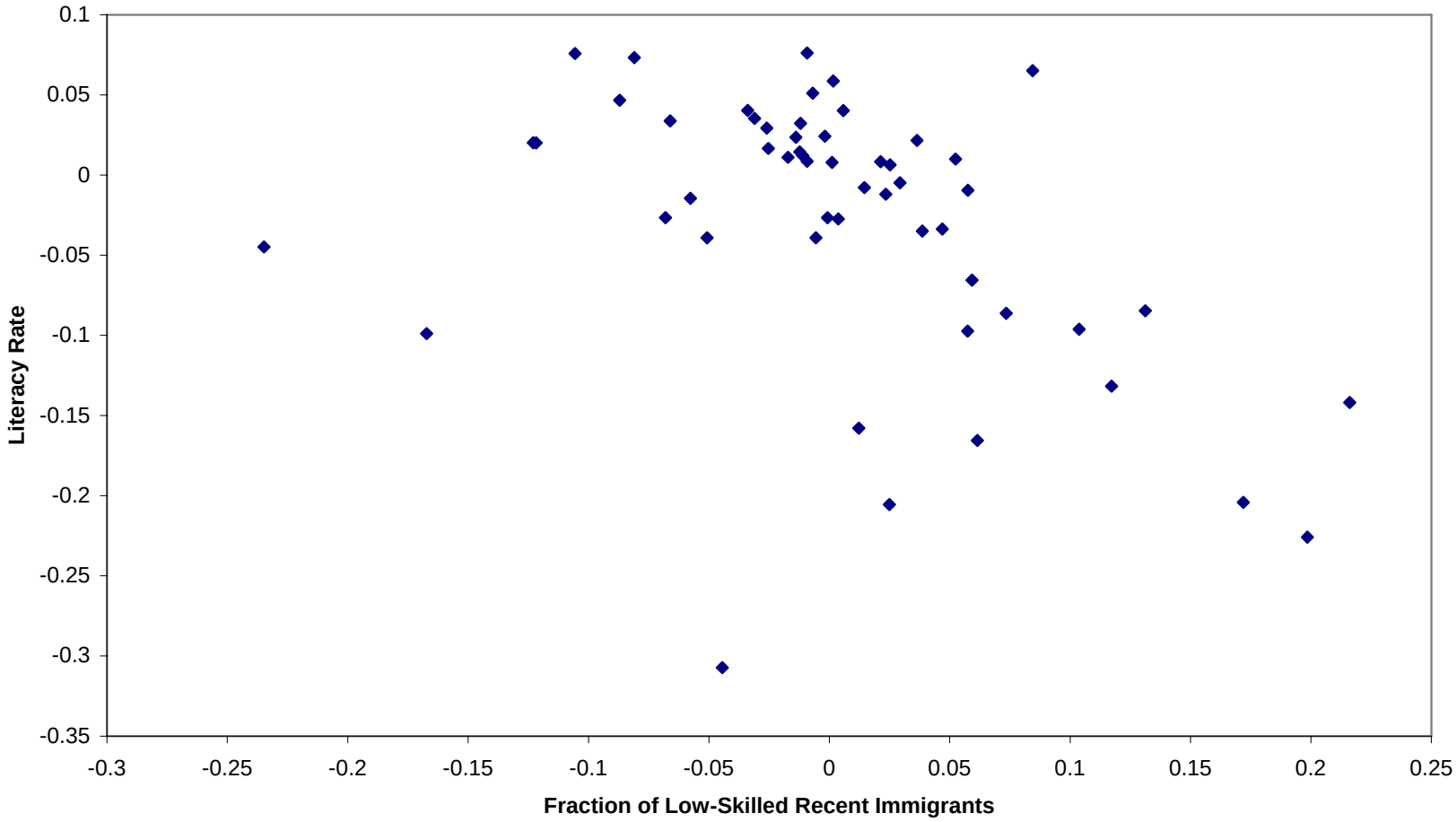

Table 1
Summary Education of Immigrants in 1910, of the Second-Generation in 1910 and 1940, and of the Third Generation in 1940, by Ethnic Group

	1910				1940			
Ethnic Group	Literacy Rate of First-Gener. Adults 30-50	Adults Sample Size	School Attend. Rate of Second-Gener. Kids 6-18	Kids Sample Size	Years of School. of Second-Gener. Adults 30-50	Adult Sample Size	School Attend. Rate of Third-Gener. Kids 6-18	Kids Sample Size
African	.754	49	.667	18	6.0	464	.853	526
Armenian	.871	39	.875	8	9.7	7	.911	31
Bulg/Ser/Cro	.699	216	.571	98	8.2	54	.890	321
Czech	.928	225	.683	498	8.8	179	.881	462
Dutch/Flemis	.944	143	.705	277	9.1	118	.856	424
English/Wels	.989	1,008	.768	2,390	10.0	874	.887	1,630
Finnish	.913	150	.692	124	8.8	113	.878	187
French/Cana	.977	135	.670	823	9.0	429	.851	929
German/Aust	.910	3,079	.690	5,307	8.9	2,940	.868	5,757
Greek	.826	150	.571	14	11.0	14	.883	298
Hungarian	.874	526	.673	165	9.4	125	.826	476
Irish	.972	1,104	.753	1,954	9.7	1,242	.904	1,912
Italian	.636	1,476	.677	1,348	8.7	902	.857	4,282
Japanese	.775	223	.721	43	10.8	8	.935	130
Jewish	.868	815	.771	1,153	10.8	472	.884	723
Lithuanian	.719	164	.686	121	8.8	71	.863	279
Mexican	.458	205	.393	201	4.1	134	.736	891
Polish	.736	338	.594	1,188	8.3	685	.846	2,184
Portuguese	.558	68	.702	114	7.5	50	.820	197
Romanian	.810	58	.727	11	11.2	37	.866	178
Russian	.748	721	.782	116	10.6	411	.882	1,791
Ruthenian	.465	43	.686	35	7.2	8	.781	58
Scandinavian	.985	1,385	.750	2,339	9.4	1,365	.896	2,402
Scottish	.992	252	.759	399	10.1	252	.917	480
Slovak	.814	250	.669	248	8.4	115	.889	219
Spanish	.886	44	.552	30	9.2	20	.891	101
Native	.910	28,031	.725	40,576	8.8	53,755	.841	60,395

Notes: The table shows the fraction of foreign-born men aged 30 to 50 who can read and write in any language in 1910, the fraction of second-generation children (i.e.: born in the US to a foreign-born parent) aged 6 to 18 who are enrolled in school in 1910, the average years of schooling for second-generation men aged 30-50 in 1940, and the fraction of third-generation children (i.e.: born in the US to a second-generation parent) who are enrolled in school in 1940. For comparison purposes, the last row shows the corresponding measures for third- and higher- generation adults (this is, US-born adults with US-born parents), and for fourth- and higher-generation children (this is, US-born children of US-parents and grandparents). *Source:* Author's tabulations from the 1910 and 1940 Census IPUMS files.

Table 2
Descriptive Statistics for the 1910 and 1920 Census IPUMS samples

Variables	1910-1920	1910	1920
A. Children (Second-Generation Americans of Schooling Age)			
	Dependent Variable		
In School	.724	.709	.737
	(.447)	(.454)	(.440)
	Covariates		
Age	10.93	11.18	10.73
	(3.98)	(4.00)	(3.95)
Female	.496	.493	.498
	(.500)	(.500)	(.500)
Number of Siblings	3.68	3.78	3.60
	(2.12)	(2.10)	(2.12)
In Metropolitan Area	.611	.567	.648
	(.487)	(.496)	(.477)
In Region:			
New England	.123	.121	.125
	(.328)	(.326)	(.330)
Middle Atlantic	.335	.296	.368
	(.472)	(.457)	(.482)
East North Central	.244	.260	.231
	(.430)	(.439)	(.421)
West North Central	.150	.180	.125
	(.357)	(.384)	(.331)
South Atlantic	.020	.018	.022
	(.141)	(.133)	(.148)
East South Central	.008	.008	.007
	(.087)	(.088)	(.085)
West South Central	.032	.034	.030
	(.175)	(.181)	(.171)
Mountain	.030	.029	.030
	(.170)	(.169)	(.171)
Pacific	.058	.054	.061
	(.233)	(.226)	(.240)
N	76,847	19,202	57,645

Notes: The data are from the Census IPUMS for 1910 and 1920. In Panels A and B, the sample is restricted to second-generation Americans (i.e.: born in the US to a foreign-born parent) of schooling age (6 to 18 years old) who reside with their parents. In Panel C, the sample is restricted to first- and second-generation Americans (i.e.: foreign-born, or born in the US to a foreign-born parent) of working age (19 to 60 years old). Standard deviations are in parentheses. All other entries are means (weighted by the IPUMS sample-line weight).

Table 2 (continued)
Descriptive Statistics for the 1910 and 1920 Census IPUMS samples

Variables	1910-1920	1910	1920
B. Fathers (First-Generation Americans with Children of Schooling Age)			
	Regressor of Interest: Parental Capital		
Literacy	.843	.853	.835
	(.364)	(.354)	(.371)
	Covariate		
Age	44.84	45.35	44.42
	(8.46)	(8.27)	(8.59)
	Instrument		
Age at Arrival	20.61	20.75	20.49
	(8.52)	(8.52)	(8.52)
N	76,847	19,202	57,645
C. Adults (First- and Second-Generation Americans of Working Age)			
	Regressor of Interest: Ethnic Capital		
Literate	.863	.832	.890
	(.344)	(.374)	(.312)
	Demographic Characteristics		
Age	36.71	35.67	37.09
	(11.29)	(11.17)	(11.33)
Fraction in Metropolitan Area	.625	.595	.653
	(.480)	(.491)	(.476)
	Instrument		
Low-Skilled New Immigrants as Fraction of Population	.043	.079	.011
	(.106)	(.146)	(.021)
New Immigrants as Fraction of Population	.082	.142	.029
	(.148)	(.196)	(.040)
N	305,842	81,649	224,193

Notes: The data are from the Census IPUMS for 1910 and 1920. In Panels A and B, the sample is restricted to second-generation Americans of schooling age (6 to 18 years old) who reside with their parents. In Panel C, the sample is restricted to first- and second-generation Americans of working age (19 to 60 years old). Standard deviations are in parentheses. All other entries are means (weighted by the IPUMS sample-line weight).

Table 3
Endogamy Rates in the First and Second Generation (1910-1920), by Ethnic Group

Ethnic Group	First-Generation Women			Second-Generation Women		
	Married Native (1)	Married Same Group (2)	Married Other Group (3)	Married Native (4)	Married Same Group (5)	Married Other Group (6)
African (Black)	25.4	60.8	13.9	71.4	22.6	5.9
Armenian	1.0	97.9	1.0	0.0	100.0	0.0
Bulg/Serb/Croa	0.2	95.8	4.0	18.0	64.1	18.0
Czech	2.6	87.5	9.9	13.6	61.1	25.3
Dutch/Flemish	6.0	83.6	10.5	27.0	45.8	27.3
English/Welsh	27.1	52.9	19.9	53.2	23.7	23.1
Finnish	0.8	92.3	6.9	10.5	62.7	26.9
French/Canadian	9.4	74.5	16.2	33.9	34.3	31.8
German/Austrian	7.2	82.9	9.9	31.0	52.3	16.6
Greek	0.4	95.5	4.1	14.3	57.1	28.6
Hungarian	0.8	87.2	12.1	11.8	41.2	47.1
Irish	11.2	71.3	17.5	33.9	39.3	26.9
Italian	0.2	98.9	0.8	7.6	82.4	10.0
Japanese	0.0	99.6	0.4	0.0	95.5	4.6
Jewish	0.2	97.8	2.0	2.8	84.2	13.0
Lithuanian	0.2	96.9	3.0	5.7	75.5	18.9
Mexican	4.2	93.5	2.3	11.7	82.8	5.5
Polish	0.5	94.3	5.3	6.4	76.1	17.5
Portuguese	1.7	95.1	3.2	14.5	64.0	21.5
Romanian	0.6	83.4	16.0	5.3	36.8	57.9
Russian	1.2	88.3	10.6	12.0	56.7	31.3
Ruthenian	0.0	88.9	11.1	0.0	71.4	28.6
Scandinavian	6.2	85.0	8.8	29.5	49.4	21.0
Scottish	20.3	47.4	32.4	51.5	9.8	38.7
Slovak	0.4	92.3	7.3	3.3	75.8	20.9
Spanish	7.1	80.1	12.8	26.7	36.0	37.3
NEC	8.0	74.1	18.0	40.2	24.0	35.8

Notes: The table shows the distribution of husband's ethnicity for married women in the 1910 and 1920 Censuses with spouse present in the household. Columns (1)-(3) show the ethnicity distribution of husbands for foreign-born women, while columns (4)-(6) do the same for second-generation women. Columns (2) and (5) refer to husbands, either first or second generation, of the same ethnic group as the wife. Columns (3) and (6) refer to husbands, either first or second generation, of some ethnic group other than that of the wife.
Source: Author's tabulations from the 1910 and 1920 Census IPUMS files.

Table 4
Descriptive Statistics for the 1910 and 1920 Census IPUMS samples, by Ethnicity

	1910					1920				
Ethnicity	Fraction Children in School	Fraction Literate Fathers	Children Sample Size	Average Literacy Rate	Adults Sample Size	Fraction Children in School	Fraction Literate Fathers	Children Sample Size	Average Literacy Rate	Adults Sample Size
African (Black)	.667	.722	18	.795	223	.687	.812	128	.852	991
Armenian	.875	.875	8	.486	91	.797	.812	69	.717	328
Bulg/Serb/Croa/Slov	.571	.724	98	.401	670	.735	.789	837	.694	1,921
Czech	.683	.783	498	.798	1,258	.717	.876	1,132	.930	3,744
Dutch/Flemish	.705	.931	277	.894	841	.723	.942	774	.959	2,858
English/Welsh	.768	.983	2,390	.988	11,812	.771	.981	4,926	.991	31,270
Finnish	.692	.871	124	.650	473	.757	.806	724	.839	1,392
French/Canadian	.67	.744	823	.845	3,269	.723	.832	1,861	.926	8,655
German/Austrian	.69	.898	5,307	.942	22,791	.714	.943	9,930	.975	56,601
Greek	.571	.643	14	.342	475	.715	.860	200	.801	1,574
Hungarian (Magyar)	.673	.848	165	.406	858	.773	.809	1,215	.762	2,867
Irish	.753	.973	1,954	.985	13,606	.75	.986	3,235	.992	30,188
Italian	.677	.602	1,348	.413	4,847	.752	.653	8,695	.722	15,987
Japanese	.721	.256	43	.422	543	.704	.577	362	.628	1,360
Jewish	.771	.864	1,153	.755	3,343	.81	.892	4,860	.901	11,188
Lithuanian	.686	.595	121	.445	489	.735	.698	859	.657	1,802
Mexican	.393	.144	201	.159	840	.405	.291	762	.317	3,749
Polish	.594	.599	1,188	.437	3,563	.684	.696	5,953	.709	12,304
Portuguese	.702	.395	114	.538	352	.685	.540	531	.624	1,301
Romanian	.727	.909	11	.510	132	.76	.901	161	.780	643
Russian	.782	.836	116	.566	430	.778	.796	1,572	.782	4,081
Ruthenian	.686	.457	35	.264	106	.648	.739	119	.657	236
Scandinavian	.75	.962	2,339	.957	6,626	.758	.974	5,217	.983	19,886
Scottish	.759	.989	399	.993	1,984	.784	.990	834	.995	5,462
Slovak	.669	.628	248	.387	753	.707	.772	1,809	.759	2,962
Spanish	.552	.567	30	.525	200	.661	.576	132	.691	798
Not Elsewhere Clas.	.752	.761	180	.513	1,074	.737	.733	748	.746	3,788

Table 5
OLS and Probit Estimates of the Effect of Parental and Ethnic Capital on Individual Skills

	Using Father's Literacy					Using average of Father's and Mother's Literacy
	(OLS) In School	(OLS) In School	(OLS) In School	(OLS) In School	(Probit) In School	(OLS) In School
	(1)	(2)	(3)	(4)	(5)	(6)
Father's Literacy (Parental Capital)	.054**	.052**	.051**	.049**	.057**	.060**
	(.011)	(.011)	(.009)	(.008)	(.010)	(.008)
Average Literacy (Ethnic Capital)	.215**	.214**	.137**	.135**	.136**	.119**
	(.034)	(.035)	(.022)	(.022)	(.021)	(.022)
Region, Metro Effects?	No	Yes	No	Yes	Yes	Yes
Age Dummies?	No	No	Yes	Yes	Yes	Yes
Father's Age and Age-Squared?	No	No	Yes	Yes	Yes	Yes
Female			.002	.001	.001	-.003
			(.005)	(.005)	(.006)	(.005)
Father's Age at Arrival			-.0012**	-.0011**	-.0013**	-.0008**
			(.0003)	(.0002)	(.0003)	(.0003)
N	69,864	69,864	69,864	69,864	69,864	56,308

Notes: Column 5 reports the estimated marginal effects from a probit model of child's school attendance on parental English literacy and on the average English literacy rate in the ethnic group; all other entries are OLS estimated coefficients from a linear regression. Standard errors corrected for ethnicity-year clustering are reported in parentheses. The data are from the Census IPUMS for 1910 and 1920, with the sample being restricted to second-generation Americans of schooling age (6 to 18 years old) who reside with their parents. Ethnicity-year-specific average literacy rates are computed from a sample restricted to first- and second-generation Americans of working age (19 to 60 years old). All models include Census year and ethnicity main effects. In column 6, the sample is restricted to children whose parents are both first-generation Americans and are both present in the household. Single (double) asterisk denotes statistical significance at the 90% (95%) level of confidence in a one-tailed test.

Table 6
OLS Estimates of the Effect of Parental and Ethnic Capital on Individual Skills: Additional Results

	Benchmark	Coefficient on Parental Capital =0	Coefficient on Parental Capital =.2	Coefficient on Ethnic Capital =.1	Coefficient on Ethnic Capital =0
	In School	In School	In School	In School	In School
	(1)	(2)	(3)	(4)	(5)
Father's Literacy (Parental Capital)	.049** (.008)	.000	.200	.049** (.009)	.048** (.009)
Average Literacy (Ethnic Capital)	.135** (.022)	.144** (.024)	.106** (.018)	.100	.000
N	69,864	69,864	69,864	69,864	69,864

Notes: Standard errors corrected for ethnicity-year clustering are reported in parentheses. The data are from the Census IPUMS for 1910 and 1920, with the sample being restricted to second-generation Americans of schooling age (6 to 18 years old) who reside with their parents. Ethnicity-year-specific average literacy rates are computed from a sample restricted to first- and second-generation Americans of working age (19 to 60 years old). All regressions include Census year, ethnicity, region and female main effects as well as father's age, father's age squared, father's age at arrival, number of siblings, a dummy indicating residence in a metropolitan area, and a vector of age dummies. Single (double) asterisk denotes statistical significance at the 90% (95%) level of confidence in a one-tailed test.

Table 7
Age at Arrival from Non-English Speaking Country as Instrument for Parental Capital: First-Stage Estimates

	Father's Literacy (1)	Father's Literacy (2)	Father's Literacy (3)	Father's Literacy (4)	Average of Father's and Mother's Literacy (5)
Father's Age at Arrival * Non-English Speaking Country of Origin	-.0065** (.0012)	-.0065** (.0012)	-.0066** (.0012)	-.0065** (.0012)	-.0053** (.0017)
Father's Age at Arrival	-.0007** (.0002)	-.0008** (.0002)	-.0007** (.0003)	-.0007** (.0003)	-.0030** (.0011)
Region, Metro Effects?	No	Yes	Yes	Yes	Yes
Age Dummies?	No	No	Yes	Yes	Yes
Father's Age and Age-Squared, Female Dummy, and Number of Siblings?	No	No	Yes	Yes	Yes
Fraction of New Immigrants in Low-Skilled Occupations				-.058 (.070)	-.119 (.096)
N	69,864	69,864	69,864	69,864	56,308

Notes: Entries are OLS estimated coefficients from a regression of parental English literacy on parental age at arrival interacted with a dummy for non-English speaking country of birth. Standard errors corrected for ethnicity-year clustering are reported in parentheses. The data are from the Census IPUMS for 1910 and 1920, with the sample being restricted to second-generation Americans of schooling age (6 to 18 years old) who reside with their parents. All regressions include Census year and ethnicity main effects. Single (double) asterisk denotes statistical significance at the 90% (95%) level of confidence in a one-tailed test.

Table 8
Recent Immigrant Flows as Instrument for Ethnic Capital:
First-Stage Estimates

	Average Literacy (1)	Average Literacy (2)	Average Literacy (3)	Average Literacy (4)	Average Literacy (5)
Fraction of New Immigrants in Low-Skilled Occupations	-.599** (.300)	-.605** (.197)	-.605** (.197)	-.603** (.197)	-.603** (.197)
Region, Metro Effects?	No	Yes	Yes	Yes	Yes
Age Dummies?	No	No	Yes	Yes	Yes
Father's Age and Age-Squared, Female Dummy, and Number of Siblings?	No	No	No	Yes	Yes
Father's Age at Arrival * Non-English Speaking Country of Origin					-.0002 (.0002)
N	54	69,864	69,864	69,864	69,864

Notes: Entries are OLS estimated coefficients from a regression of average English literacy among first- and second-generation adults in the ethnic group on the fraction of recently arrived low-skilled immigrants. Standard errors, corrected for ethnicity-year clustering in Columns 2 through 5, are reported in parentheses. The data are from the Census IPUMS for 1910 and 1920, with the sample being restricted to first- and second-generation Americans of working age (19 to 60 years old). The fraction of new (i.e.: in the 5 years prior to the Census year) immigrant arrivals who were laborers or servants or agricultural workers, by ethnicity and year, are computed from the immigration records in Ferenczi and Willcox (1929). The mean fraction of new immigrants who were laborers/servants across ethnicity-year cells is .483, and the interquartile range (from .312 to .683) is .371. (See Appendix for more details). All regressions include Census year and ethnicity main effects. Single (double) asterisk denotes statistical significance at the 90% (95%) level of confidence in a one-tailed test.

Table 9
2SLS and NLIV Estimates of the Effect of Parental and Ethnic Capital on Individual Skills

	Using Father's Literacy				Using average of Father's and Mother's Literacy
	(2SLS) In School	(2SLS) In School	(2SLS) In School	(IV Probit) In School	(2SLS) In School
	(1)	(2)	(3)	(4)	(5)
Father's Literacy (Parental Capital)	.280**	.281**	.203**	.265	.149**
	(.078)	(.080)	(.089)	(.270)	(.069)
Average Literacy (Ethnic Capital)	.138**	.138**	.116**	.101*	.094*
	(.066)	(.066)	(.041)	(.054)	(.051)
Region, Metro Effects?	No	Yes	Yes	Yes	Yes
Age Dummies?	No	No	Yes	Yes	Yes
Father's Age and Age-Squared?	No	No	Yes	Yes	Yes
Female			.002	.001	-003
			(.005)	(.006)	(.005)
Father's Age at Arrival			-.0003	-.0004	-.0001
			(.0005)	(.0011)	(.0007)
N	69,864	69,864	69,864	69,864	56,308

Notes: Entries are IV (2SLS) estimates from a regression of child's school attendance on parental English literacy and on the average English literacy rate in the ethnic group, except for Column 4, which reports the estimated marginal effects from a IV-probit model of child's school attendance (where parental and average ethnic literacy are treated as endogenous) estimated by maximum likelihood. The excluded instruments are the father's age at arrival interacted with a dummy for non-English speaking country of origin, and the fraction of recently arrived low-skilled immigrants by ethnicity and year. Standard errors corrected for ethnicity-year clustering are reported in parentheses. The data are from the Census IPUMS for 1910 and 1920, with the sample being restricted to second-generation Americans of schooling age (6 to 18 years old) who reside with their parents. Ethnicity-year-specific average literacy rates are computed as in Table 8. All models include Census year and ethnicity main effects, a female dummy, father's age at arrival, number of siblings, a dummy indicating residence in a metropolitan area, and a full set of age dummies. Single (double) asterisk denotes statistical significance at the 90% (95%) level of confidence in a one-tailed test.

Table 10
2SLS Estimates of the Effect of Parental and Ethnic Capital on Individual Skills: Additional Results

	Benchmark	Coefficient on Parental Capital =0	Coefficient on Parental Capital =.2	Coefficient on Ethnic Capital =.1	Coefficient on Ethnic Capital =0
	In School	In School	In School	In School	In School
	(1)	(2)	(3)	(4)	(5)
Father's Literacy (Parental Capital)	.203** (.089)	.000	.200	.204** (.088)	.211** (.087)
Average Literacy (Ethnic Capital)	.116** (.041)	.153** (.045)	.117** (.035)	.100	.000
N	69,864	69,864	69,864	69,864	69,864

Notes: Standard errors corrected for ethnicity-year clustering are reported in parentheses. The data are from the Census IPUMS for 1910 and 1920, with the sample being restricted to second-generation Americans of schooling age (6 to 18 years old) who reside with their parents. The excluded instruments are the father's age at arrival interacted with a dummy for non-English speaking country of origin, and the fraction of new immigrant arrivals, in the 5 years prior to the Census year, who were laborers or servants or agricultural workers, by ethnicity and year. Ethnicity-year-specific average literacy rates are computed from a sample restricted to first- and second-generation Americans of working age (19 to 60 years old). All regressions include Census year, ethnicity, region and female main effects as well as father's age, father's age squared, father's age at arrival, number of siblings, and a vector of age dummies. Single (double) asterisk denotes statistical significance at the 90% (95%) level of confidence in a one-tailed test.

Table 11
OLS and 2SLS Estimates of the Effect of Parental and Ethnic Capital on Individual Skills: Specification Checks

	Including Father's Occupation		Excluding Children Aged 17-18		Excluding Five Smallest Ethnic Groups	
	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Father's Literacy (Parental Capital)	.047**	.198**	.047**	.211**	.050**	.205**
	(.008)	(.093)	(.009)	(.081)	(.009)	(.091)
Average Literacy (Ethnic Capital)	.147**	.122**	.129**	.121**	.138**	.116**
	(.021)	(.042)	(.020)	(.036)	(.023)	(.043)
Father is in a Low-Skilled Occupation	-.016**	-.008				
	(.004)	(.006)				
N	69,864	69,864	61,755	61,755	69,266	69,266

Notes: Standard errors corrected for ethnicity-year clustering are reported in parentheses. The data are from the Census IPUMS for 1910 and 1920, with the sample being restricted to second-generation Americans of schooling age (6 to 18 years old) who reside with their parents, except in Columns 5 and 6, where the sample is further restricted to children aged 6 to 16 years only. The excluded instruments are the father's age at arrival interacted with a dummy for non-English speaking country of origin, and the fraction of new immigrant arrivals, in the 5 years prior to the Census year, who were laborers or servants or agricultural workers, by ethnicity and year. Ethnicity-year-specific average literacy rates are computed from a sample restricted to first- and second-generation Americans of working age (19 to 60 years old). The low-skilled occupations used in Columns 1 and 2 are agriculture, laborers and servants (the same ones used in the construction of the instrument for ethnic capital). The five ethnic groups excluded in Columns 3 and 4 are African, Spanish, Romanian, Armenian and Ruthenian, and correspond to the 5 rows with the smallest counts in Table 2. All regressions include Census year, ethnicity, region and female main effects as well as father's age, father's age squared, father's age at arrival, number of siblings, and a vector of age dummies. Single (double) asterisk denotes statistical significance at the 90% (95%) level of confidence in a one-tailed test.

Table 12
OLS and 2SLS Estimates of the Effect of Parental and Ethnic Capital on Individual Skills: Exploring the Role of Geographic Concentration

	OLS (1)	OLS (2)	2SLS (3)	2SLS (4)
Father's Literacy (Parental Capital)	.049**		.203**	
	(.008)		(.089)	
Father's Literacy (Parental Capital) * High Concentration		.047**		.231
		(.013)		(.142)
Father's Literacy (Parental Capital) * Low Concentration		.048**		.182**
		(.009)		(.076)
Average Literacy (Ethnic Capital)	.135**		.116**	
	(.022)		(.041)	
Average Literacy (Ethnic Capital) * High Concentration		.261*		.141
		(.157)		(.195)
Average Literacy (Ethnic Capital) * Low Concentration		.122**		.109
		(.030)		(.084)
N	69,864	69,864	69,864	69,864

Notes: Standard errors corrected for ethnicity-year clustering are reported in parentheses. The data are from the Census IPUMS for 1910 and 1920, with the sample being restricted to second-generation Americans of schooling age (6 to 18 years old) who reside with their parents. The concentration index is computed as the fraction of all adults of working age in the region who are first- or second- generation and who have the same ethnicity, and averages approximately .12 for the entire sample. High (Low) concentration is then defined as a dummy that equals one if the individual lives in a region where their ethnic group (first and second generation) comprises 12% or more (less than 12%) of the population of working age, zero otherwise. The excluded instruments are the father's age at arrival interacted with a dummy for non-English speaking country of origin, and the fraction of new immigrant arrivals, in the 5 years prior to the Census year, who were laborers or servants or agricultural workers, by ethnicity and year. Ethnicity-year-specific average literacy rates are computed from a sample restricted to first- and second-generation Americans of working age (19 to 60 years old). All regressions include Census year and ethnicity main effects, a female dummy, father's age at arrival, number of siblings, a dummy indicating residence in a metropolitan area, and a full set of age dummies. Single (double) asterisk denotes statistical significance at the 90% (95%) level of confidence in a one-tailed test.

Table 13

OLS and 2SLS Estimates of the Effect of Parental and Ethnic Capital on Individual Skills: Using Regional Endogamy Rates as a Measure of the Level of Interaction Within Groups

	OLS	OLS	2SLS	2SLS
	(1)	(2)	(3)	(4)
Father's Literacy (Parental Capital)	.049**		.203**	
	(.008)		(.089)	
Father's Literacy (Parental Capital) * High Endogamy Rate		.041**		.171**
		(.011)		(.075)
Father's Literacy (Parental Capital) * Low Endogamy Rate		.062**		.259**
		(.006)		(.118)
Average Literacy (Ethnic Capital)	.135**		.116**	
	(.022)		(.041)	
Average Literacy (Ethnic Capital) * High Endogamy Rate		.141**		.140**
		(.023)		(.040)
Average Literacy (Ethnic Capital) * Low Endogamy Rate		.095**		.017
		(.028)		(.082)
N	69,864	69,864	69,864	69,864

Notes: Standard errors corrected for ethnicity-year clustering are reported in parentheses. The data are from the Census IPUMS for 1910 and 1920, with the sample being restricted to second-generation Americans of schooling age (6 to 18 years old) who reside with their parents. The endogamy rate is computed as the fraction of married women in the region whose husband belongs to the same ethnic group, and averages approximately .55 for the entire sample. High (Low) Endogamy Rate is then defined as a dummy that equals one if the individual lives in a region where the endogamy rate for their ethnic group equals 55% or more (less than 55%), zero otherwise. The excluded instruments are the father's age at arrival interacted with a dummy for non-English speaking country of origin, and the fraction of new immigrant arrivals, in the 5 years prior to the Census year, who were laborers or servants or agricultural workers, by ethnicity and year. Ethnicity-year-specific average literacy rates are computed from a sample restricted to first- and second-generation Americans of working age (19 to 60 years old). All regressions include Census year and ethnicity main effects, a female dummy, father's age at arrival, number of siblings, a dummy indicating residence in a metropolitan area, and a full set of age dummies. Single (double) asterisk denotes statistical significance at the 90% (95%) level of confidence in a one-tailed test.