# The Effects of Spanish-Language Background on Completed Schooling and Aptitude Test Scores *

Luis Locay† Tracy L. Regan‡ Arthur M. Diamond, Jr.§

**Abstract**

We investigate the effect of speaking Spanish, for Hispanic children who grew up in the U.S., on completed schooling and aptitude test scores using the NLSY79. We model the accumulation of traditional human capital and English fluency, leading to the joint determination of the outcome variables. We find that speaking Spanish reduces test scores with no significant effect on completed schooling. The reduced test scores: 1) increase in three of the tests when the parents are more educated; 2) are more dramatic when the choice of home language is endogenous; and 3) are not systematically greater for verbal tests.

**JEL Classification:** I20, J15.

**Keywords:** Hispanics, language, aptitude, ability, NLSY79.

*The authors are grateful for comments received at the 2007 American Economic Association Meetings, with special thanks to Aimee Chin, the 2006 Southern Economic Association $79^{th}$ Annual Conference, the 2006 LACEA-LAMES Conference, the 2006 Guanajuato Workshop for Young Economists, the 2006 Annual Congress of the European Society for Population Economics, the 2006 Western Economic Association International $81^{st}$ Annual Conference, and from participants in the applied microeconomics/labor workshop at the University of Miami. We thank Sara Michalski, Isaac Petit-Frere, and Debanjali Roy for research assistance.

†Department of Economics, University of Miami, P.O. Box 248126, Coral Gables, FL 33124-6550; Phone: (305) 284-1502; Fax: (305) 284-2985; Email: llocay@bus.miami.edu.

‡Corresponding author: Department of Economics, University of Miami, P.O. Box 248126, Coral Gables, FL 33124-6550; Phone: (305) 284-1654; Fax: (305) 284-2985; Email: tregan@miami.edu.

§Department of Economics, University of Nebraska at Omaha, 6001 Dodge St., Omaha, NE 68182-0048; Phone: (402) 554-3647; Fax: (402) 554-2853; Email: adiamond@mail.unomaha.edu.

# 1 Introduction

Does growing up in the U.S. in a home where Spanish is spoken affect an individual's cognitive abilities as measured by aptitude tests? Does it have any influence on how much schooling that individual eventually acquires? The purpose of this paper is to investigate these questions using data on persons of Hispanic ancestry who grew up in the U.S. While there is an extensive literature linking earnings to aptitude test scores and educational attainment (two areas in which Hispanics continue to lag non-Hispanic whites), not much attention has been devoted to whether these measures of intellectual development are themselves affected by the language spoken at home. As the Hispanic share of the population in the U.S. continues to grow, understanding the determinants of Hispanic test scores and educational attainment is of increasing importance.

According to the Current Population Survey, there were 40.4 million Hispanics living in the U.S. in 2004—14.0 percent of the population, which represents an increase from 12.6 percent in 2000. The differences in educational attainment between Hispanics and non-Hispanic whites are quite stark. In 2005, 25.1 percent of Hispanics, age 25 and above, had less than nine years of education, while only 3.3 percent of non-Hispanic whites did. At higher education levels, only 12.1 percent of Hispanics had a bachelor's degree or more, whereas the corresponding figure for non-Hispanic whites was 30.6 percent.

Differences in standardized test scores are also large. In 2004, for example, the average verbal and math SAT scores for Hispanics were 456.3 and 458.3, respectively. Correspondingly, the averages for non-Hispanic whites were 528 and 531.[1] In the two verbal Armed Services Vocational Aptitude Battery (ASVAB) tests used in this paper, scores for Hispanics and non-Hispanic whites are 20.5 and 26.7 on the word knowledge exam and 8.6 and 11.2 on the paragraph comprehension exam. Similar differences arise for the two math ASVAB tests we used: 13.2 and 18.4 points for the arithmetic reasoning exam and 9.9 and 13.8 points for the math knowledge exam.[2]

Related topics have been explored by various authors. For example, McManus (1990) finds that the return to English proficiency is lower in Spanish enclaves. Chiswick (1991) and Gonzalez (2000) find greater returns to speaking fluency than to reading fluency; the latter also finds a premium to writing skills over reading skills. Recently, Bleakley and Chin (2004, 2008) have addressed the possible correlation between measures of English proficiency and the error term in wage regressions by exploiting the cognitive theory that children learn languages more easily at younger ages. This hypothesis is also explored by Chiswick and Miller (2008), Chiswick et al. (2005), and Gonzalez (2003). For the families of Hispanic children born or raised in the U.S., however, the issue is not so much whether or not their children should become fluent

[1] See www.collegeboard.com.

[2] At the time, the maximum verbal and math SAT score were 800. The corresponding maximum ASVAB word knowledge, paragraph comprehension, arithmetic reasoning, and math knowledge scores are 35, 15, 30, and 25. The numbers mentioned here do not match the figures presented in Table 1 as they correspond to all the Hispanics and non-Hispanic whites contained in the data set, and not to the subset used in our empirical analysis.

in English—as they overwhelmingly do—but whether they should, to the extent that they are able, expose them to Spanish. Being bilingual has obvious benefits, but being raised in a home where Spanish is spoken may have drawbacks in an English-speaking society.

To our knowledge, the economics literature has been mostly silent on the effect of speaking Spanish at home on educational attainment and aptitude test scores. Fryer and Levitt (2006) mention in passing that speaking Spanish at home has little effect on the initial gap or the trajectory of test scores between Hispanics and non-Hispanic whites, but they provide no formal results, as this was not the focus of their study. Other papers investigating ethnic and racial differences in test scores include Clotfelter et al. (2009) and Fryer and Levitt (2004).[3] Rosenthal et al. (1983), from the sociology literature, use a nationally representative sample of elementary students and find a negative relationship between speaking Spanish at home and verbal and math aptitude, with the effect being stronger for the former than for the latter.

In Section 2 of this paper we construct a formal model of the joint determination of schooling level and aptitude test scores. We view tests scores as measures of an individual's human capital at the time the test is administered. We modify Ben-Porath's (1967) model of human capital to allow for two types of complementary capital: traditional human capital (measured by aptitude tests) and English fluency. Speaking Spanish at home can slow down the acquisition of English fluency, thus impairing the transmission of skills, resulting in lower test scores. How much it does so may depend on the characteristics of the parents.[4] The implications of the model for the effect of speaking Spanish at home on schooling levels are ambiguous, however. Anything that lowers the productivity of the process of human capital accumulation reduces both the marginal benefits and the marginal costs of staying in school.

Section 3 presents the empirical formulation we employ. A detailed description of the data is given in Section 4 and in the Data Appendix. We use data from the National Longitudinal Survey of Youth 1979 (NLSY79). The NLSY79 has been widely used in the labor economics literature, but to our knowledge, this paper is the first to use it in exploring the relationship between Spanish language background and academic achievement. The nature of the NLSY79 enables us to determine a respondent's final level of schooling, and it also contains verbal and math aptitude test scores from the ASVAB tests. Section 5 presents our results. We find that speaking Spanish at home as a child reduces tests scores, but has no statistically significant effect on schooling levels. As much as 19-34 percent of the White-Hispanic test differential can be accounted for by speaking Spanish at home. We also find that for three of the tests the reduction in test scores from speaking Spanish at home tend to increase with parents' schooling.

Section 5 treats speaking Spanish at home as a child as an exogenous characteristic of families. In Section 6 we model the decision to speak Spanish at home, and we find that endogenizing it substantially

[3]On a somewhat related note, Angrist et al. (2008) study the effect of using English as the medium of instruction on English proficiency in Puerto Rico.

[4]We will use 'skills,' 'knowledge,' and 'human capital' interchangeably throughout the paper.

increases its negative effect on test scores. Doing so helps to explain an even greater fraction of the White-Hispanic test differential (59-97 percent). We conclude in Section 7 with a summary of our results.

# 2 Conceptual Framework and Empirical Specification

Like Hansen et al. (2004) we assume that aptitude test scores are not measures of innate ability, but rather outcome variables that themselves are generated in part by this latent ability. We view the arithmetic reasoning, math knowledge, word knowledge, and paragraph comprehension test scores as measures of different types of human capital at the time the test is administered. In what follows, however, we take a simplified approach and treat human capital as all of one type.

Our model is a modified version of Ben-Porath's (1967) model of human capital. In the period $[0, t_s]$, which we refer to as the schooling/home period, the individual is born, lives with his parents, and attends school but does not work. During this time the individual invests in human capital and becomes fluent in English. By English fluency we mean the ability to understand basic spoken English, and do not mean the full mastery of verbal skills. According to this interpretation of fluency, the average high school student who has spoken only English his entire life is considered as fluent as a Shakespearian scholar. At time $t_s$ the individual enters the labor force and uses his acquired human capital to generate income and for further investment.

## 2.1 The Schooling/Home Period

During the schooling/home period human capital is accumulated according to the following:

$$\dot{h} = bh^{\alpha}\left(\frac{E}{2}\right)^{\eta},\ 1 > \alpha,\ \eta > 0, \tag{1}$$
$$\frac{\dot{E}}{E} = g > 0 \text{ if } E < 2,$$
$$\dot{E} = 0 \text{ otherwise},$$
$$E(0) = 1,\ h(0) = h_0,$$

where $h(t)$ is human capital and $E(t)$ is fluency in English, both at time $t$, which coincides with the individual's age. Fluency is acquired at an exogenous rate, $g$, determined by one's language environment, including that of the home. Initial human capital, $h_0$, and the productivity parameter of the human capital accumulation equation, $b$, may also be related to observable household characteristics. Everyone begins life with a level of English fluency, $E(0) = 1$, and achieves fluency at time $t_F$ where $E(t_F) = 2$. The time at which fluency is achieved is given by $e^{gt_F} = 2$, or $t_F = \frac{ln(2)}{g}$. We assume that everyone achieves fluency before the end of the schooling period, i.e. $t_F < t_s$. (In the empirical work we assume that fluency has been achieved by the time the aptitude test is administered, which for some individuals is before they have completed their schooling.)

The model summarized in equation (1) does not allow for any positive influences from the knowledge of Spanish in terms of human capital accumulation. The model could be easily modified to allow for such benefits of bilingualism if the empirical work to follow suggests that such a modification would be fruitful. For the period $[0, t_F]$ the solution to the differential equation for human capital in equation (1) is

$$h(t) = \left[ h_0^{1-\alpha} + \frac{b(1-\alpha)}{\eta g 2^{\eta}} (e^{\eta g t} - 1) \right]^{\frac{1}{(1-\alpha)}}. \tag{2}$$

From equation (2) we can obtain $h(t_F)$, which is the initial condition for the human capital differential equation for the period between achieving fluency and completing school, i.e. $[t_F, t_s]$. For the period $[t_F, t_s]$ human capital is given by:

$$h(t) = \left[ h_0^{1-\alpha} + b(1-\alpha) \left( t - \frac{\pi}{\eta g} \right) \right]^{\frac{1}{(1-\alpha)}}, \tag{3}$$

where $\pi = 2^{-\eta} + \eta ln(2) - 1 > 0$ and $t > \frac{\pi}{\eta g}$. From equation (3) we can obtain $h(t_s)$, which is the initial condition for human capital during the working period, i.e. $[t_s, T]$, where $T$ is the time of retirement.

In the empirical implementation, we assume that every member of our sample was administered the aptitude test after they achieved fluency. This implies that for a person who was in school at the time of the test, $t_\tau$, equation (3) gives that person's level of human capital, and consequently his test score, $\tau$. Define $s_\tau$ as the years of schooling at the time the test is administered. Then $s_\tau = t_\tau - 6$, and we can express the person's test score as:[5]

$$\tau(s_\tau, g, b, h_0) = h(s_\tau + 6). \tag{4}$$

We can determine from equations (3) and (4) that:

$$\frac{\partial \tau}{\partial s_\tau} = b(h(t_\tau))^{\alpha} > 0, \tag{5}$$

$$\frac{\partial \tau}{\partial g} = \frac{b\pi(h(t_\tau))^{\alpha}}{\eta g^2} > 0,$$

$$\frac{\partial \tau}{\partial b} = \frac{(h(\tau))^{\alpha}(\eta g t_\tau - \pi)}{\eta g} > \frac{(h(t_\tau))^{\alpha}(1 - 2^{-\eta})}{\eta g} > 0,$$

$$\frac{\partial \tau}{\partial h_0} = \frac{(h(t_\tau))^{\alpha}}{h_0^{\alpha}} > 0.$$

[5]This assumes that an individual begins his schooling at age six—the typical age in the U.S.

The first condition in equation (5) states that the longer the individual is in school before he takes the exam, the higher his human capital at the time of the test and therefore the higher the test score. The next three conditions state that an increase in the rate of accumulation of English fluency, $g$, in the productivity of human capital accumulation, $b$, or in the initial human capital endowment, $h_0$, will raise the test score at the time the test is administered.

## 2.2 The Working Period

After completing school the individual enters the labor market with human capital $h_s = h(t_s)$ given by equation (3). Following Ben-Porath (1967) we assume that the individual can allocate any portion of his human capital to generate earnings, $y(t)$, or to generate more human capital through on-the-job training. Let $x(t)$ be the amount of human capital devoted to on-the-job training at time $t$. The individual's earnings are then given by:

$$y(t) = w(h(t) - x(t)), \tag{6}$$

where $w$ is the wage rate per unit of human capital. We assume that the production of human capital through on-the-job training is governed by a process similar to that of the schooling/home period:

$$\dot{h} = ax^{\alpha}, \tag{7}$$

where $a$ is a productivity parameter. The individual's objective is to determine the path of investment in human capital, $x(t)$, that maximizes the present value of earnings, $y(t)$, minus the direct cost of schooling, which is given by:

$$\int_{t_s}^{T} e^{-rt} y(t)dt - p \int_{0}^{t_s} e^{-rt}dt = \int_{t_s}^{T} e^{-rt} y(t)dt - \frac{p(1 - e^{-rt_s})}{r}, \tag{8}$$

where $p$ is the direct cost of schooling per unit of time and $r$ is the discount rate. Taking $t_s$ as given, equation (8) is maximized subject to equations (6) and (7), the boundary condition $h(s) = h_s$, and the non-negativity constraint, $x(t) \geq 0$.

At the start of the working period, $t_s$, the individual will invest positive amounts in his human capital. That investment will decline over time as retirement age approaches. At time $T$ investment in human capital becomes zero. Let us define potential work experience, $t_w$, as $t_w = t - t_s$. The solution to the working period problem, which is $x(t) = \left(\frac{a\alpha}{r}(1 - e^{r(t-T)})\right)^{\frac{1}{(1-\alpha)}}$, gives rise to the optimal path of human capital given by:

$$h(t) = h(t_w + t_s) = h_s + \int_{0}^{t_w} a(x(z + t_s))^{\alpha} dz. \tag{9}$$

The first term on the right hand side of (9) is the human capital at the end of the schooling period, and the second term on the right hand side is the net accumulation of human capital between the end of the schooling period, $t_s$, and time $t$.

Equation (9) provides the human capital for someone who has left school by the time the test is administered. Since for such a person the years of schooling at the time of the test, $s_\tau$, equals the years of completed schooling, $s$, we can express his test score as:

$$\tau(s, t_w, a, g, b, h_0) = h(t), \tag{10}$$

where $h(t)$ is given by equation (9).

According to equation (9), human capital at the end of the schooling period is reflected one-for-one in human capital during the working period. The effects of $(s, g, b, h_0)$ on the test score after the schooling period are therefore the same as those during the schooling period, and are given by equation (5). From equation (9) we can derive the effect of $t_w$ and $a$ on the test score as:

$$\frac{\partial \tau}{\partial t_w} = a(x(t_w + t_s))^\alpha > 0 \text{ if } a > 0, \tag{11}$$
$$\frac{\partial \tau}{\partial a} = \int_0^{t_w} \frac{a(x(z + t_s))^\alpha}{1 - \alpha} dz > 0 \text{ if } a > 0.$$

The optimal value of the present value of income net of the direct schooling costs is

$$V(h_s, t_s) = \frac{w(e^{-rt_s} - e^{-rT})h_s}{r} + w \int_{t_s}^{T} e^{-rt} \int_{t_s}^{t} a x^\alpha(z) dz dt - w \int_{t_s}^{T} e^{-rt} x(t) dt \tag{12}$$
$$- \frac{p(1 - e^{-rt_s})}{r}.$$

The first term on the right hand side of equation (12) is the present value of earnings generated by the human capital acquired during the schooling/home period. The next term is the present value of the earnings generated by the human capital accumulated during the working life. From the first two terms, the third term, representing the present value of foregone earnings from on-the-job investment in human capital, must be subtracted. The fourth term on the right hand side is the present value of direct schooling costs.

From equation (12) we obtain that:

$$V_{h_s} = \frac{w(e^{-rt_s} - e^{-rT})}{r} > 0. \tag{13}$$

Increasing human capital at the end of the schooling period by one unit permanently raises the path of human capital by one unit. This in turn increases earnings by $w$ per unit of time. Equation (13) gives the present value of that increase in earnings. Similarly for changing the date of completion of schooling:

$$V_{t_s} = -w\left[h_s e^{-rt_s} + \frac{a(x(t_s))^{\alpha}(e^{-rt_s} - e^{-rT})}{r}\right] + we^{-rt_s}x(t_s) - pe^{-rt_s} < 0. \tag{14}$$

Holding human capital at the end of the schooling/home period, $h_s$, constant, extending the period of schooling increases the direct cost of schooling and reduces the period of potential earnings. The latter is partially offset by a reduction in post-schooling investment, but the overall effect is to reduce the present value of earnings.

The individual's problem is then to choose the length of the schooling/home period so as to maximize equation (12), where $h(s)$ is given by equation (3). Since $t_s = s + 6$, we can express the objective function as $v(s; a, g, b, h_0) = V(h(s+6), s+6)$, where $h(t)$ is given by equation (3). The first order condition for this problem is given by the following:

$$v_s(s; a, g, b, h_0) = V_{h_s} b(h(t_s))^{\alpha} + V_{t_s} = 0, \tag{15}$$

where $V_{h_s}$ is given by equation (13) and $V_{t_s}$ by equation (14). The solution to equation (15) is the optimal level of schooling, $s(a, g, b, h_0)$. The effects of $(a, g, b, h_0)$ on schooling can be obtained by differentiating equation (15). Their effect, however, will generally be ambiguous. Consider, for example, the effect of more rapid growth of English fluency. Differentiating equation (15), we obtain:

$$\frac{\partial s(a, g, b, h_0)}{\partial g} = -\frac{v_{sg}}{v_{ss}}. \tag{16}$$

The second order conditions require that $v_{ss} < 0$, so the sign of $\frac{\partial s}{\partial g}$ will be the same as that of $v_{sg}$. The latter is given by:

$$v_{sg} = \frac{wb\pi(h(t_s))^{\alpha}}{\eta g^2}\left[\frac{b\alpha(e^{-rt_s} - e^{-rT})(h(t_s))^{\alpha-1}}{r} - e^{-rt_s}\right], \tag{17}$$

which cannot be signed. A higher rate of growth of English fluency makes the schooling/home period more productive in the acquisition of human capital, encouraging the individual to stay in school longer. Since this results in higher human capital at every point in time, it results in greater foregone earnings from staying in school. Our model predicts that the overall effect on completed schooling is thus ambiguous.

## 2.3 Speaking Spanish at Home and Parental Schooling

Our main interest in this paper is to assess the effects of speaking Spanish at home as a child on aptitude test scores, and secondarily, because the theoretical predictions are ambiguous, on schooling. For now we will treat speaking Spanish at home as an exogenous variable. In Section 6 we endogenize the decision of the language environment of the home.

Of the three productivity parameters in the schooling/home period of the model developed above, $(g, b, h_0)$, the most natural one through which speaking Spanish at home should influence the acquisition of human capital is the growth rate of English fluency, $g$. We expect that a child who grows up in a home where the parents are fluent in English, but nevertheless speak Spanish, will become fluent in English more slowly. As shown above, this will result in lower test scores, but it will have an ambiguous effect on the level of schooling. In a household where the parents are not fluent in English, the alternative to speaking Spanish may be to speak very poor English, which may lower rather than raise the rate at which a child becomes fluent in English. Presumably such a household would be a Spanish speaking one. The effect of speaking Spanish at home on the accumulation of human capital, therefore, may interact in complex ways with parental characteristics.

Unfortunately, our data set does not include measures of the English fluency of the parents. It does, however, include their levels of schooling and whether they were born abroad, both of which we believe are correlated with their degree of fluency in English. Furthermore, parental schooling should be directly related to the other productivity parameters, namely $b$ and $h_0$, and through these indirectly related in the production of human capital to speaking Spanish at home.

We can envision at least four ways in which parental schooling can impact the accumulation of human capital and therefore the test scores of the individuals in our sample. First, parental schooling is likely to be correlated with the innate ability of the parents and thus with their children's inherited abilities. In our model, higher innate ability can manifest itself in terms of higher values of $h_0$, $b$, or $g$. Second, parents with higher levels of education are likely to have higher incomes, which will be associated with better quality schools and other inputs into the human capital accumulation process. Furthermore, parents with more schooling are likely to have more knowledge to impart to their children and be better at doing so. These last two should work through increasing the productivity parameters $b$ and $g$. Finally, better educated parents are more likely to be fluent in English and this can affect the growth rate of English directly as well as through its interaction with the language spoken at home, as stated above.

Perhaps the potential interaction between ability and home language that first springs to mind is that if speaking Spanish at home reduces $g$, the rate at which fluency is acquired, its negative effects will be weaker for more able individuals. If this effect is present, the test scores of the children of more educated parents should not be reduced by as much by speaking Spanish at home as those of less educated parents. There are, however, other ways in which ability and home language can interact. From equation (5) it can be shown that $\frac{\partial^2 \tau}{\partial g \partial h_0} > 0$ and $\frac{\partial^2 \tau}{\partial g \partial b} > 0$. This means that any reduction in $g$ resulting from speaking Spanish at home will have a larger negative effect on test scores the greater are $h_0$ and $b$. The more productive parents are at imparting human capital to their children, the greater the opportunity cost for parents fluent in English to speak Spanish at home. According to this effect, speaking Spanish at home will reduce test scores more for children of more educated parents.

## 3 The Empirical Implementation

The model from the previous section gives rise to the following two equations for completed schooling and test scores:

$$s(a, b, g, h_0) \tag{18}$$

$$\tau_i(s_\tau, t_w, a, b, g, h_0). \tag{19}$$

For our empirical implementation we will assume that the schooling equation (18) is given by:

$$s = \gamma Z + u, \tag{20}$$

where $Z$ is a set of observable determinants of schooling and $u$ is an error term, which we assume is distributed as $N(0, \sigma_u)$. Among the elements included in $Z$ are individual and family characteristics, including parents' schooling, and measures of the direct and indirect costs of schooling similar to those used in Hansen et al. (2004).

We assume for an individual's score on test $i$, $\tau_i$, the following empirical formulation:

$$\tau_i = f_i(s_p)\psi + \theta_i s_\tau + \lambda_i t_w + \beta_i X + \varepsilon_i \text{ for } i = 1, ..., 4, \tag{21}$$

where $\psi$ takes on the value of "1" if an individual spoke Spanish at home as a child, and "0" otherwise, $s_p$ is a vector of the schooling level of each parent, $f_i()$ is a function of $s_p$, $X$ is a vector of individual and family characteristics (again including measures of parental schooling), and $\varepsilon_i$ is the error term representing unobservable determinants of the test score.

Equation (21) cannot be estimated by OLS because $s_\tau$ is likely to be correlated with the error term. The ASVAB tests were administered to all the participants in the NLSY79 in the same year, 1980. Respondents, therefore, varied in age at the time they took the aptitude tests. If more able respondents also obtain more schooling, then the more able are more likely to be in school when the tests were administered. Years of schooling in 1980 would in part measure the effects of unobserved ability. To correct for this bias we substitute the expected value of $s_\tau$ for its actual value in estimating equation (21).

Define $\Delta = t_\tau - 6$ as the maximum years of schooling an individual could have completed at the time of the test.[6] Let $I = 1$ if $s \leq \Delta$, and $I = 0$ otherwise. If $I = 0$ an individual is still in school at the time of

[6] Again this assumes that individuals begin their schooling at the typical age of six. We did, however, adjust $\Delta$ for whether

the test and so $s_\tau = \Delta$ and $t_w = 0$. If $I = 1$ an individual has completed his schooling at the time of the test and so $s_\tau = s$ and $t_w = t_\tau - s - 6 = \Delta - s$. We can now restate equation (21) as:

$$\tau_i = \begin{cases} f_i(s_p)\psi + \theta_i \Delta + \beta_i X + \varepsilon_i \text{ if } I = 0 \\ f_i(s_p)\psi + \theta_i \gamma Z + \lambda_i(\Delta - \gamma Z) + \beta_i X + (\theta_i - \lambda_i)u + \varepsilon_i \text{ if } I = 1. \end{cases} \tag{22}$$

From equation (22) we can compute the expected score of test $i$. Let $c = \frac{(\Delta - \gamma Z)}{\sigma_u}$, and $\Phi(c)$ and $\phi(c)$ be the standard normal distribution and density evaluated at $c$. The expected score of test $i$ is given by:

$$E(\tau_i) = f_i(s_p)\psi + \theta_i[\Delta(1 - \Phi(c)) + \gamma Z \Phi(c) - \sigma_u \phi(c)] + \lambda_i[(\Delta - \gamma Z)\Phi(c) + \sigma_u \phi(c)] + \beta_i X. \tag{23}$$

We will employ a two-step estimation procedure. From equation (20) we obtain that:

$$Prob(I = 1) = Prob(s \leq \Delta) = Prob\left(\frac{u}{\sigma_u} \leq \frac{(\Delta - \gamma Z)}{\sigma_u}\right) = \Phi(c), \tag{24}$$

$$Prob(I = 0) = 1 - \Phi(c).$$

The first-stage of our procedure is to estimate the probit given by equation (24) to obtain $\widehat{\gamma}$ and $\widehat{\sigma}_u$. We use these estimates to construct $\Phi(\widehat{c})$ and $\phi(\widehat{c})$ and then substitute them into equation (23) and estimate it by OLS. This procedure provides us with consistent coefficient estimates, and the standard errors are corrected using a bootstrap technique.

## 4 Data

The data used in this study are from the National Longitudinal Survey of Youth 1979 (NLSY79). Much of the early work on earnings and English proficiency relied on Census data or the 1976 Survey of Income and Education (SIE)—the former because of its large sample size and the latter because of its richness of language-related questions. To our knowledge, the NLSY79 has not been used in economics to study the effects of language background. The singular exception to this is Bleakley and Chin (2008) who use it as a robustness check to the results obtained using data from the Census.[7] We focus our attention on a set of Hispanics who were either born in the U.S. or migrated here before they were age seven.[8] Our set

an individual began school earlier or later than age six.

[7] Other possible data sets include the Early Childhood Longitudinal Study of Kindergartners (ECLS-K), the National Education Longitudinal Study (NELS-88), the Education Longitudinal Study (ELS-02), and the National Adult Literacy Survey (NALS). While these data sets may offer some advantages such as better measures of English fluency, they also have some drawbacks. Among these are that the individuals involved may be very young, or that completed schooling is not observed, or that there is no variability in the age at which respondents were administered the aptitude tests, thus not allowing us to estimate structural test score equations. Nevertheless, in future work it may be valuable to explore such data sets.

[8] This latter condition eliminates any complications that may arise from an individual receiving part or all of his education abroad, and from the decision of whether to become fluent in English. For the purposes of this paper, we consider Puerto Rico to be a "foreign" country because English is not the primary language.

of Hispanics include Cubans, Mexicans (i.e. Chicanos, Mexicans, and Mexican-Americans), Puerto Ricans, and other Hispanics (i.e. other Hispanics and other Spanish).

The two outcome variables of interest in our analysis are the completed level of schooling and the ASVAB test scores. We consider four of the ASVAB tests which assess an individual's math and verbal skills: 1) arithmetic reasoning; 2) math knowledge; 3) word knowledge; and 4) paragraph comprehension.[9] The primary variable of interest in our empirical implementation is whether a respondent spoke Spanish at home as a child. We constructed this variable by first noting if an individual indicated speaking a language other than English at home as a child, and then if this language was Spanish. The other key explanatory variables of interest are parental schooling levels. The Data Appendix explains the construction of the data in greater detail and discusses other variables and data sets used in the analysis.

The final sample we use in the empirical strategy is comprised of 1,312 Hispanics—612 males and 700 females. This sample omits non-Hispanics along with Hispanics who migrated to the U.S. after age six, as well as any other individuals who were missing information for the relevant variables. A subsample of 624 Hispanics for whom both parents were born in the U.S. was also used for some of the analysis.

Table 1 provides descriptive statistics for some of the key variables for the selected sample. Descriptive statistics for the entire set of variables are contained in Appendix Table A.1. Hispanic males (females) comprise 46.6 (53.4) percent of the overall sample. The majority (65 percent) of the sample is Mexican. The second largest ethnicity is Puerto Rican (19.4 percent). Cubans are the smallest group represented (5.4 percent). About 86 percent of the sample reports speaking Spanish at home as a child. On average, both parents have completed a little more than eight years of schooling. On average, the individuals themselves report completing 12.7 years of schooling; 11.2 years of which were completed at the time the ASVAB tests were administered. Those individuals who report speaking Spanish at home as a child complete slightly less school (12.7 years) than those who do not speak Spanish at home (13.1 years), and about the same as those whose parents were born in the U.S. Similarly for the ASVAB tests, those who indicate speaking Spanish at home score lower than those who do not. Specifically, there is a 11.3, 9.6, 13.7, and 10.7 percentage point differential for the arithmetic reasoning, math knowledge, word knowledge, and paragraph comprehension tests between the non-Spanish and Spanish speakers. Test scores for those whose parents were both born in the U.S. were between those who spoke Spanish at home and those who spoke only English, but they were closer to the former than the latter. In terms of parental schooling, the parents of those who did not speak Spanish at home average three more years of schooling than those who did. Once again, those whose parents were borsht U.S.-born had average parental schooling in between those of the two language groups.

For purposes of comparison, Table 1 also provides information on a selected sample of non-Hispanic whites (for sake of brevity, we will refer to this group as 'Whites'). This group is comprised of individuals

[9]Since 1989 the Department of Defense has used these four tests in constructing a percentile score for the overall Armed Forces Qualification Test (AFQT) and they are also the only four tests used by Hansen et al. (2004).

who: 1) are classified as white by the interviewer; 2) report no Hispanic ancestry; 3) are born in the U.S. and whose parents were born in the U.S. as well; 4) report only speaking English at home as a child; and 5) lived in a county that included at least one respondent from our Hispanic sample. Whites score higher on all four tests, and have more completed schooling (as do their parents), than not only Hispanics, in general, but also Hispanics who did not speak Spanish at home.

# 5 Estimation and Results

The entire set of estimated parameters for the completed schooling equation (20) obtained from the first-stage estimation of equation (24), can be found in Appendix Table B.1, column 1. Our discussion will be limited to the primary variables of interest, namely, speaking Spanish at home as a child and levels of parental schooling. Table 2 presents the coefficient estimates for speaking Spanish at home and parental schooling. Speaking Spanish at home reduces completed schooling, but the effect is not statistically significant. An additional year of maternal schooling increases the respondent's completed schooling by only 0.04 years, and it is also not statistically significant. The effects are quite different for father's schooling. For every additional year of paternal schooling, the respondent's completed schooling rises by 0.16 years, and is statistically significant at the one percent level.[10]

The entire second-stage results corresponding to equation (23) are contained in Appendix Table B.2. For each aptitude test we estimated two versions of equation (23). In the first version, we set $f_i(s_p) = \alpha_{i0}$, implying that any effect of speaking Spanish at home on test scores would be independent of parental schooling levels.[11] This version was estimated on the entire sample and on a subsample consisting of respondents for whom both parents were born in the U.S. In the second version we allow the effect of speaking Spanish at home to vary by parental schooling levels.[12] For reasons given below, this version was estimated only for the subsample of individuals for whom both parents were born in the U.S. Highlights of the results are presented in Table 3. The dependent variable in each regression is the standardized test score.[13] For each test the first column presents the selected results for the entire sample when the effect of speaking Spanish at home does not vary by parental schooling level. For the entire sample more schooling at the time the test is administered results in higher test scores for all four tests. An additional year of schooling increases test scores by 0.076-0.118 standard deviations. For potential work experience, however, the results for the math

[10]The effect of parents' schooling in Table 2 refers to the schooling of parents who were born in the U.S. and who were present when the respondent was age 14. Table 2 also shows mother's and father's schooling are also interacted with whether they were born abroad. The interaction of mother's and father's schooling and whether they were absent when the respondent was age 14 can be found in Appendix Table B.1, column 1.

[11]The corresponding first-stage probit, whose estimated coefficients are shown in column 1 of Appendix Table B.1, also does not allow for the effect of speaking Spanish at home to vary with parental schooling levels.

[12]The corresponding first-stage probit results are shown in column 3 of Appendix Table B.2 that allows for the coefficient on speaking Spanish at home to vary by parental schooling levels.

[13]The test scores are standardized by the overall mean and standard deviations for the combined Hispanic and White sample.

tests are markedly different than for the verbal tests. An additional year of post-schooling experience has little effect on the math scores, whereas it increases the word knowledge and paragraph comprehension test scores by about 0.09 and 0.08 standard deviations, respectively, and both are statistically significant at the one percent level. These results imply that math skills for Hispanics stop improving with the completion of formal schooling, while verbal skills continue to improve at nearly the same pace after entering the labor force as during the schooling/home period.

For parental schooling, we first consider the results when the effects of speaking Spanish at home do not vary with the level of parents' education (see Table 3, columns 1, 4, 7, and 10). The effect of mother's schooling is statistically significant and similar in magnitude (0.03-0.06) across all four tests. Similarly for father's schooling, the coefficients vary little (0.05-0.06) across all four tests and are always statistically significant.[14]

Turning to the variable of most interest, we see from Table 3 that, speaking Spanish at home reduces test scores, with the effects being statistically significant at conventional levels for the math tests—arithmetic reasoning and math knowledge—and for word knowledge, but statistically insignificant for paragraph comprehension.[15] To give some sense of the magnitude of these effects, speaking Spanish at home reduces test scores by more than a reduction of two years of schooling for the arithmetic reasoning, math knowledge, and word knowledge tests and a little over one year for the paragraph comprehension test. To check whether our results were due to the way we treated endogenous schooling in the second stage regressions, we estimated the same test equations substituting the respondent's age at the time the tests were taken for the expected years of schooling and potential work experience. The results are shown in Table B.4 in the appendix. As can be seen, the estimated coefficients on speaking Spanish at home are extremely close to those reported in Table 3 for each of the four tests.

We turn now to investigating the interaction of parental schooling levels with speaking Spanish at home. We believe that during the relevant period individuals of high ability living abroad would have been more likely to be hindered in obtaining education by financial considerations than comparable persons in the U.S. It seems likely, therefore, that schooling is a better indicator of ability for those individuals educated in the U.S. than those educated abroad. For this reason, when looking at the interaction of parental schooling with language spoken at home, we restrict the analysis to individuals whose parents were both born in the U.S.[16]

Solely for purposes of comparison with the results using the entire sample, we estimated both completed schooling and test score equations for the subsample of respondents for whom both parents were born in the

[14]By comparison, Currie and Thomas (1999) using a different data set find a larger effect of maternal schooling than paternal schooling on children's test scores.

[15]For paragraph comprehension the coefficient on speaking Spanish at home as a child is statistically significant at the 11 percent level.

[16]We do not know where parents were educated but it is reasonable to assume that if the parents were born in the U.S. they were likely educated here. Parents born elsewhere could have been educated either in the U.S. or abroad.

U.S. The complete first-stage estimates, under the assumption that the effects of speaking Spanish at home are independent of parents' schooling levels, appear in column 2 of appendix Table B.1, and selected results in column 2 of Table 2. As can be seen, the results are similar for both samples. The complete second-stage results appear in columns 2, 5, 8, and 11 of appendix Table B.2, and selected results in Table 3. Most of the coefficients reported in the respective columns of Table 3 estimated using the subsample are similar in magnitude to their counterparts estimated using the entire sample.[17] The exception is years of schooling, which appears to be less important for the subsample of respondents whose parents were born in the U.S. than for the entire sample.

To investigate how the effect of speaking Spanish at home varies with parents' schooling we let $f_i(s_p)$ in equation (21) take on distinct values depending on whether the average of both parents' schooling levels $s_p$ is strictly less than nine years, greater than or equal to nine but strictly less than 12 years, and greater than or equal to 12 years.[18] The complete first-stage results from this estimation using the subsmaple of respondents with U.S.-born parents appears in columns 3 of Table B.1, in the appendix, and selected results in column 3 of Table 2. For the second-stage the complete results appear in columns 3, 6, 9, and 12 of Table B.2, in the appendix. The same columns in Table 3 show the selected results. For arithmetic reasoning, math knowledge, and paragraph comprehension, the negative effect on test scores of speaking Spanish at home increases with parents' education. The increase is more pronounced for the two math tests.[19] For word knowledge the effect of speaking Spanish at home as a child is similar for all levels of parents' schooling, though the point estimate is smallest for the highest parental schooling category.

We can also assess the magnitude of the effects of speaking Spanish at home by seeing how much of the White-Hispanic test score differentials it explains. We perform this comparison by first estimating the equivalent of equation (23) for our sample of Whites, and use the results to calculate Blinder-Oaxaca type decompositions.[20] For each test let $\varphi_i$ represent the parameters of test score equation (21), and let $\kappa$ represent the parameters of the probit for the test equation (24).[21] Following the analysis in Bauer and Sinning (2008), we decompose the White-Hispanic test score differentials as follows:

$$\begin{aligned} & E_{\varphi_{w_i},\kappa_w}(\tau_{w_i}|Z_w, X_w, \Delta_w) - E_{\varphi_{h_i},\kappa_h}(\tau_{h_i}|Z_h, X_h, \Delta_h, s_p, \psi) = \\ & \left[E_{\varphi_{w_i},\kappa_w}(\tau_{w_i}|Z_w, X_w, \Delta_w) - E_{\varphi_{w_i},\kappa_h}(\tau_{h_i}|Z_h, X_h, \Delta_h, s_p, \psi)\right] + \\ & \left[E_{\varphi_{w_i},\kappa_h}(\tau_{h_i}|Z_h, X_h, \Delta_h, s_p, \psi) - E_{\varphi_{h_i},\kappa_h}(\tau_{h_i}|Z_h, X_h, \Delta_h, s_p, \psi)\right] \end{aligned} \quad (25)$$

[17]Not surprisingly, statistical significance levels are sometimes lower when the subsample is used.

[18]Note that we used alternative definitions of parental schooling levels (i.e. minimum, maximum) and the results are virtually unchanged.

[19]It is also the case that the differences betwen the two higher parental schooling levels and the lowest one are statistically significant only for the math tests.

[20]A description of the procedure used for estimating equation (23) for Whites appears in the Technical Appendix. The only difference from the approach used for Hispanics is that for Whites we correct for censoring.

[21]$\kappa$ also includes the parameters from the probits used to correct for censoring for Whites. See the Technical Appendix for further details.

where the subscripts $w$ and $h$ stand for White and Hispanic, respectively, and the subscripts on the expectation operator show the parameters used when computing an expectation. The first term on the right hand side of equation (25) is the difference due to covariates, and the second term is the difference due to parameters.

Table 4 shows the decomposition of White-Hispanic standardized test score differentials into differences due to covariates and parameters. The first decomposition is based on the model that does not allow the effect of speaking Spanish at home to vary with parental schooling levels. As can be seen differences in covariates account for most of the White advantage in test scores. Most of the difference due to covariates is in turn accounted for by parental schooling. Differences in parameters are small and for two tests—math knowledge and paragraph comprehension—actually favor Hispanics. On all four tests, speaking Spanish at home handicaps Hispanics from 0.11 to 0.23 standard deviations. It accounts for 34 percent of the White-Hispanic differential for arithmetic reasoning, 33 percent for word knowledge, 27 percent for math knowledge, and 19 percent for paragraph comprehension. Parameter differences other than speaking Spanish at home actually favor Hispanics on all four tests. The second decomposition shown in Table 4 is based on coefficients estimated using the subsample of respondents whose parents were both born in the U.S. The most notable change from the previous decomposition is that parental schooling accounts for a smaller share of the difference due to covariates than when the entire sample is used. This is to be expected, as the schooling levels of U.S.-born parents is on average higher. The third panel is Table 3 shows the decompositions based on estimates derived from the subsample, and allowing the effect of speaking Spanish to vary by parental schooling levels. As can be seen, for the math tests the importance of speaking Spanish at home is reduced. The reason for this is a combination of a negative effect of speaking Spanish at home that increases in magnitude with the level of parental schooling for both math tests, and schooling parameter differences that are evaluated at Hispanic means, which for parental schooling are much lower for Hispanics than for Whites.[22]

# 6 Endogenous Choice of Home Language

So far we have treated whether a family speaks Spanish at home as exogenous. The concern is that the main variable of interest may be related to some unobserved characteristic of families that negatively impacts test scores, and that we have misattributed its effects to speaking Spanish at home. It is possible, for example, that families who in the 1960s and 1970s were more concerned with the education of their children, tended to speak only English at home. If this was the case, then the effect of this unobserved, heightened concern for education would be incorrectly attributed to the language spoken at home.[23] The reverse, of

[22] This is the well-known index problem with the Blinder-Oaxaca decomposition method.

[23] This raises the question of why such families wished to avoid speaking Spanish in the first place. Did they believe that doing so would handicap their children in some way? It also suggests that speaking Spanish at home would have a stronger

course, is just as plausible. It may well be that families that emphasize education are also more likely to value their children being able to speak a second language, or have stronger loyalty to their ancestral culture. If this was the case, then we may well have underestimated the negative effects on test scores of speaking Spanish at home. In this section we develop and implement a model where speaking Spanish at home is endogenous.

Without altering the human capital acquisition process of our model, if families are going to speak Spanish at home even when it reduces their children's human capital, it must be because they place some value on doing so. Let $v(s, \psi; Z_h)$ be the net present value of income introduced in Section 2.2, where $Z_h$ is a vector of variables other than $\psi$ that can affect the acquisition of human capital. In that section we treated $\psi$ ($\psi = 1$ if Spanish was spoken at home) as exogenous. Suppose now that the family values income and speaking Spanish. It is then interested in choosing $s$ and $\psi$ so as to maximize a more general utility function:

$$U(\psi, v(s, \psi; Z_h); Z_s), \tag{26}$$

where $Z_s$ is a vector of variables that affects a family's tastes for speaking Spanish at home. The solution to the maximization of equation (26) is a pair of equations, $s(Z)$ and $\psi(Z)$, where $Z = (Z_h, Z_s)$.

For the empirical implementation, we maintain as much of the previous structure as possible. Essentially, this means that $\psi$ is replaced with $P[\psi = 1]$ in equation (23). Let us assume that:

$$\psi(Z) = \begin{cases} 1, \text{ if } \delta Z + u_s \geq 0 \\ 0, \text{ if } \delta Z + u_s < 0. \end{cases} \tag{27}$$

Then the probability Spanish is spoken at home is

$$P[\psi(Z) = 1] = P[u_s \geq -\delta Z] = 1 - \Phi(-\delta Z), \tag{28}$$

where we have assumed that $u_s \sim N(0, 1)$. The schooling equation, which takes on the same form as before, is

$$s(Z) = \gamma Z + u_h. \tag{29}$$

Thus, the probability that an individual has completed his schooling at the time the test is given remains:

$$P[s(Z) \leq \Delta] = P\left[\frac{u_h}{\sigma_h} \leq \frac{(\Delta - \gamma Z)}{\sigma_h}\right] = \Phi(c). \tag{30}$$

effect on years of completed schooling than on test scores which is not what we found in the previous section.

The expected test score for test $i$ is now given by:

$$E(\tau_i) = f_i(s_p)[1 - \Phi(-\delta Z)] + \theta_i[\Delta(1 - \Phi(c)) + \gamma Z\Phi(c)] + \lambda_i[(\Delta - \gamma Z)\Phi(c) + \sigma_h\phi(c)] + \beta_i X. \qquad (31)$$

Our first-stage procedure now consists of estimating equations (28) and (30) as a bivariate probit to obtain $\widehat{\delta}$, $\widehat{\gamma}$, and $\widehat{\sigma}_h$. We then use these to construct $\phi(\widehat{c})$ and $\Phi(\widehat{c})$ and substitute them into equation (31), which we estimate by OLS. As before, the standard errors are corrected using a bootstrap technique.

In order to prevent our identification from relying solely on functional forms, we need to have some variables that affect a family's taste for speaking Spanish at home but do not directly affect the test scores. That is, there need to be some variables in $Z_s$ that are not included in $Z_h$. We use three variables to this end. The first is the percent of Hispanics residing in the county where the respondent lived at age 17. The idea is that families in which the parents have poor English skills or put more value on their children growing up in areas with a greater Hispanic influence, are more likely to settle in communities with a large Hispanic population. A concern with excluding the percent Hispanic variable from the test score equations is that it may be correlated with community attributes that directly affect test scores. It may be the case, for example, that Hispanics tend to live in poorer communities that have lower quality schools. To investigate this possibility, we included the percent Hispanic in the test score regressions for our sample of Whites. We found that the coefficient on percent Hispanic was virtually zero for arithmetic reasoning, math knowledge and paragraph comprehension, and it was positive and significant at the 10 percent level for word knowledge. It does not appear to be the case, therefore, that percent Hispanic is a proxy for lower quality of schools or for some other community attribute that would directly lower test scores. A second variable we use is one that we believe captures how closely an individual identifies with his Hispanic ancestry. We have termed this measure "Hispanicity." The NLSY79 allowed respondents to list up to six ethnic identities. For this study we classified an individual as Hispanic if he listed a Hispanic ethnicity among the first four. Individuals reporting only Hispanic ethnicities, or reporting them earlier on the list, were considered more Hispanic. For example, an individual who only reports being Mexican is considered "more" Hispanic than an individual who first reports being Irish and then indicates being Mexican. It is less clear than with percent Hispanic variable how the Hispanicity variable would directly affect the test scores, but perhaps it is correlated with family resources, even after controlling for other covariates. We investigated this possibility by regressing the log of family income on the indicator variables "moderately Hispanic" and "more Hispanic," and the family characteristic variables included in the test score equations, using a sample of Hispanics who were ages 16 or younger at the start of the NLSY79.[24] The coefficients on the two Hispanicity variables were small,

[24] When the NLSY79 began in 1979, the respondents were between ages 14 and 22. Consequently, many respondents could have been living on their own at this point and so questions referring to an individual's family are somewhat ambiguous. In our attempts to ensure that the family income measures corresponded to individuals who were still living with their families,

statistically insignificant, and not even of consistent sign. Finally, for the third variable we constructed an indicator variable that takes on the value of "1" if either the mother or father was born abroad, "0" otherwise. We believe that whether Spanish is spoken at home depends more on the level of fluency of the parent with the poorest English skills, as opposed to the average level of fluency of the parents. Presumably, parents who are foreign-born are more likely to have poorer English skills. Since we are already controlling for foreign birth of the parents in the test score equations, we saw no reason to be concerned about this new indicator variable having a direct effect on test scores.[25] For more detail on the construction of these variables, please refer to the Data Appendix.

Since we are now treating speaking Spanish at home as an endogenous variable, the first-stage results of the bivariate probit are of less interest. The complete results can be seen in columns 4-7 of Appendix Table B.1. Columns 4 and 6 show the coefficients for the schooling and probability of speaking Spanish at home equations, respectively, estimated using the entire sample. Columns 5 and 7 show the corresponding results when the subsample of respondents whose parents were born in the U.S. is used. We simply point out here that in the probability of speaking Spanish at home equation (28), the percent Hispanic in the county of residence at age 17 and the measures of the degree of Hispanicity are of the expected sign and statistically significant at the one percent level.

Table 5 presents selected independent variables of interest for equation (31); the entire second-stage results can be found in Appendix Table B.4. Once again, we estimated two versions of the aptitude test score equations: one where we did not allow the effect of speaking Spanish at home to vary by parental schooling levels, and one where it did. The first version was estimated using both the entire sample (columns 1, 4, 7, and 10 in Tables 5 and B.4), and the subsample of respondents whose parents were both born in the U.S. (columns 2, 5, 8, and 11 in Tables 5 and B.4). The second version was estimated only on the subsample. For variables not involving the probability of speaking Spanish at home as a child, the estimated coefficients in Table 5 are quite similar in magnitude and statistical significance to the corresponding estimates in Table 3, where the choice of home language was treated as exogenous. Treating home language as endogenous has a substantial effect on the coefficient estimates for those variables that include speaking Spanish at home as a child, however. A comparison of the estimates in Table 5 with those in Table 3 shows that treating home language as endogenous amplifies the effect on test scores of speaking Spanish at home.

When we use the entire sample and do not allow the effect of speaking Spanish at home to vary by parents' schooling levels, the coefficients on speaking Spanish at home in Table 5 are from two to over three times the magnitude of their counterparts in Table 3, and they are all statistically significant at conventional levels.[26] Speaking Spanish at home is now estimated to reduce test scores by the equivalent of 4.3 to 6.2 fewer

we focused our attention on individuals ages 16 and younger. This exercise greatly reduces the sample size, however.

[25] Note also that equation (28) omits the male interactions with absent father and number of siblings. The decision to speak Spanish at home depends on the characteristics of the family, and not necessarily on a child's sex.

[26] In Table 3 the estimated coefficient on speaking Spanish at home for paragraph comprehension was only statistically

years of schooling.[27] When the subsample is used, the coefficient estimates for speaking Spanish at home as a child are similar in magnitude to those estimated on the entire sample, though statistical significance levels tend to be lower. Table 6 presents the Blinder-Oaxaca decompositions. As can be seen, speaking Spanish at home accounts for somewhere between 59 and 97 percent of the White-Hispanic differential when the entire sample is used.[28]

When we allow the effect of speaking Spanish at home to vary with parents' schooling, we also obtain that speaking Spanish at home has a stronger negative effect on all four test scores and for all three parental schooling levels then when we treated speaking Spanish as exogenous. For arithmetic reasoning, math knowledge and paragraph comprehension, we find once again that the negative effects of speaking Spanish at home rise with the level of parents' schooling, but the increase is more moderate than when home language was treated as exogenous.[29] For word knowledge, the effect of speaking Spanish at home continues to show no trend with respect to parental schooling levels.

As can be seen from the Blinder-Oaxaca decompositions in the third panel of Table 6 for arithmetic reasoning, math knowledge, and paragraph comprehension, speaking Spanish at home has a smaller effect on tests scores when the effect of that variable is allowed to vary with parents' schooling, then when it is not. The reason for this is once again that those are the three tests for which the negative effect of speaking Spanish at home increases in magnitude with parental schooling, and parameter differences are weighted by the lower average education of Hispanic parents.

# 7 Summary and Conclusions

In this paper we have provided a formal model of the accumulation of traditional human capital and English fluency which leads to the joint determination of aptitude test scores and years of schooling. In our model, speaking Spanish at home when parents are capable of speaking English will slow down the acquisition of English fluency and traditional human capital. This will result in lower aptitude test scores that we interpret as measures of various types of human capital. The implications for completed schooling levels are ambiguous, since lowering the rate at which an individual can accumulate human capital reduces both the marginal benefits and costs of remaining in school.

Our primary empirical findings are consistent with our theoretical model. When treating the choice of home language as exogenous we find negative and statistically significant effects of speaking Spanish at

significant at the 11 percent level.

[27]For arithmetic reasoning, 6.2; for math knowledge, 4.7; for word knowledge, 4.5; and for paragraph comprehension, 4.3 years.

[28]For arithmetic reasoning, 97 percent; for math knowledge, 62 percent; for word knowledge, 65 percent; and for paragraph comprehension, 62 percent of the gap is explained.

[29]The difference between the second highest parental education category and the lowest is statistically significant at the five percent level for arithmetic reasoning and at the 10 percent level for math knowledge. For arithmetic reasoning, the difference between the highest and lowest parental schooling levels is also statistically at the 10 percent level.

home for three of the four test scores: arithmetic reasoning, math knowledge, and word knowledge. The magnitude of the effects are equivalent to a reduction of about two years of schooling, and can account for between 19 and 34 percent of the gap between Whites and Hispanics in these three test scores. We also find that the magnitude of the effect increases with parents' schooling for three of the tests. This result is consistent with the notion that more educated parents face higher opportunity costs of speaking Spanish at home. In fact, a recent article in the *New York Times* highlights research by Bialystok (2001) who points out the costs associated with bilingualism (e.g., rapid access to words).[30] Bialystok (2001) is quoted as saying, "[Bilingualism] doesn't make kids smarter. There are documented cognitive developments but whatever smarter means, it isn't true."

When we treat choice of home language as endogenous, the estimated effects of speaking Spanish at home remain negative, increase substantially in magnitude, and are statistically significant at conventional levels for all four aptitude tests. This is consistent with the notion that those families that were particularly concerned with their children's education and skills were more likely to speak Spanish at home. Consequently, treating the choice of home language as exogenous understates the negative effects of speaking Spanish at home. Modeling the choice of home language as endogenous, we find that speaking Spanish at home reduces test scores by the equivalent of between 4.3 and 6.2 years of schooling, and can account for between 59 (math knowledge) and 97 (arithmetic reasoning) percent of the gap in test scores between Whites and Hispanics.

While not the primary focus of this paper, we found some interesting similarities and differences between the math and verbal tests. Contrary to our initial expectations, speaking Spanish at home does not seem to have a uniformly larger effect on either the math or verbal test scores.[31] In other respects the impact of home language on test scores differs noticeably. First, for Hispanics the accumulation of math skills seems to end with formal schooling, while the rate of accumulation of verbal skills appears to continue at nearly the same pace after entering the labor force.[32] Second, the magnitude of the reduction in test scores from speaking Spanish at home rises more sharply with parents' schooling for math than for verbal tests. We believe these two findings are consistent with one other. They suggests that the acquisition of verbal skills is less dependent than the acquisition of math skills on formal education, which can be more deeply affected by speaking Spanish at home.

All empirical work is constrained by data limitations, and ours is no different. We are particularly concerned that the NLSY79 does not contain any direct measures of the English fluency of the parents or the relevant household income. The problem with the latter arises because some individuals first appear in the NLSY79 when they are already living independently of their parents. The household incomes of such

[30] See J. Anderson's *New York Times* article, "Looking for Baby Sitters: Foreign Language a Must," published on August 18, 2010.

[31] We suspect most people, like ourselves, would have expected a stronger effect for the verbal than for the math tests.

[32] Or alternatively, perhaps for most occupations, the returns to learning additional math are much lower than the returns to learning additional English.

individuals are not comparable to those still living with their parents. This is the main reason we chose not to use household income as an explanatory variable. Furthermore, our measure of Spanish-language background is rather crude. For example, we do not know how intensively Spanish was spoken at home and by whom, nor the fluency in English and Spanish of the parents. Of less concern, at least to us, is that the NLSY79 is an older data set, and it is conceivable that the relationship between home language, schooling, and test scores among Hispanics has changed over time. The remedy for these shortcomings may be to use one of the more recent data sets mentioned in Section 5. As we stated previously, those data sets have their own drawbacks but they would at least allow us to determine if our results are due to the shortcomings of the NLSY79. We leave this for future work.

Finally, in this paper we have not addressed the interesting question of the effect of home language on labor markets. Even if speaking Spanish at home reduces aptitude test scores, as we have found in this paper, it still may be the case that there is a positive return in terms of higher wages for bilingual individuals. This is a question for which the NLSY79 is well-suited and one which we also leave for future investigation.

## References


[1] Angrist, J., A. Chin, and R. Godoy, 2008, "Is Spanish-Only Schooling Responsible for the Puerto Rican Language Gap?" *Journal of Development Economics*, 85, pp. 105-128.

[2] Antecol, H. and K. Bedard, 2002, "The Relative Earnings of Young Mexican, Black, and White Women," *Industrial and Labor Relations Review*, 56, pp. 122-135.

[3] Antecol, H. and K. Bedard, 2004, "The Racial Wage Gap: The Importance of Labor Force Attachment Differences across Black, Mexican, and White Men," *Journal of Human Resources*, 39, pp. 564-583.

[4] Aughinbaugh, A., C.R. Pierret, and D.S. Rothstein, 2005, "The Impact of Family Structure Transitions on Youth Achievement: Evidence from the Children on the NLSY79," *Demography*, 42, pp. 447-468.

[5] Bauer, T.K. and M. Sinning, 2005, "Blinder-Oaxaca Decomposition for Tobit Models," *STATA Journal*, 8, pp. 480-492.

[6] Ben-Porath, Y., 1967, "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 75(4), pp. 352-365.

[7] Bialystok, E., 2001, *Bilingualism in Development: Language, Literacy, and Cognition*. Port Chester, NY; Cambridge University Press.

[8] Bleakley, H. and A. Chin, 2004, "Language Skills and Earnings: Evidence from Childhood Immigrants," *Review of Economics and Statistics*, 86, pp. 481-496.

[9] Bleakley, H. and A. Chin, 2008, "What Holds Back the Second Generation? The Intergenerational Transmission of Language Human Capital Among Immigrants," *Journal of Human Resources*, 43, pp. 267-298.

[10] Cameron, S.V. and J.J. Heckman, 2001, "The Dynamics of Educational Attainment for Black, Hispanic, and White Males," *Journal of Political Economy*, 109, pp. 455-499.

[11] Chiswick, B.R., 1991, "Speaking, Reading, and Earnings among Low-Skilled Immigrants," *Journal of Labor Economics*, 9, pp. 149-170.

[12] Chiswick, B.R. and P.W. Miller, 2008, "The Critical Period Hypothesis for Language Learning" *Journal of Multingual and Multicultural Developments*, 29(1), pp. 16-29.

[13] Chiswick, B.R., Y.L. Lee, and P.W. Miller, 2005, "Parents and Children Talk: English Proficiency within Immigrant Families" *Review of Economics of the Household*, 3, pp. 243-268.

[14] Clotfelter, C.T., H.F. Ladd, and J.L. Vigdor, 2009, "The Academic Achievement Gap in Grades 3 to 8," *Review of Economics and Statistics*, 91(2), pp. 398-419.

[15] Currie, J. and D. Thomas, 1999, "The Intergenerational Transmission of 'Intelligence': Down the Slippery Slopes of *The Bell Curve*," *Industrial Relations*, 38, pp. 297-330.

[16] Flouri, E. and A. Buchanan, 2004, "Early Fathers and Mothers Involvement and Childs Later Educational Outcomes," *British Journal of Educational Psychology*, 74, pp. 141-153.

[17] Fryer Jr., R.G. and S.D. Levitt, 2004, "Understanding the Black-White Test Score Gap in the First Two Years of School," *The Review of Economics and Statistics*, 86, pp. 447-464.

[18] Fryer Jr., R.G. and S.D. Levitt, 2006, "The Black-White Test Score Gap Through Third Grade," *American Law and Economics Review*, pp. 1-33.

[19] Gonzalez, A., 2000, "The Acquisition and Labor Market Value of Four English Skills: New Evidence from NALS," *Contemporary Economic Policy*, 18, pp. 259-269.

[20] Gonzalez, A., 2003, "The Education and Wages of Immigrant Children: The Impact of Age at Arrival," *Economics of Education Review*, 22, pp. 203-212.

[21] Hansen, K.T., J.J. Heckman, and K.J. Mullen, 2004, "The Effect of Schooling and Ability on Achievement Test Scores," *Journal of Econometrics*, 121, pp. 39-98.

[22] Kossoudji, S.A., 1988, "English Language Ability and the Labor Market Opportunities of Hispanic and East Asian Immigrant Men," *Journal of Labor Economics*, 6, pp. 205-228.

[23] McManus, W.S., 1985, "Labor Market Costs of Language Disparity: An Interpretation of Hispanic Earnings Differences," *American Economic Review*, 75, pp. 818-827.

[24] McManus, W.S., 1990, "Labor Market Effects of Language Enclaves–Hispanic Men in the United States," *Journal of Human Resources*, 25, pp. 228-252.

[25] McManus, W., W. Gould, and F. Welch, 1983, "Earnings of Hispanic Men: The Role of English Language Proficiency," *Journal of Labor Economics*, 1, pp. 101-130.

[26] Reimers, C.W., 1983, "Labor Market Discrimination Against Hispanic and Black Men," *Review of Economics and Statistics*, 65, pp. 570-579.

[27] Rosenthal, A.S., K. Baker, and A. Ginsburg, 1983, "The Effect of Language Background on Achievement Level and Learning among Elementary School Student," *Sociology of Education*, 56, pp. 157-169.

[28] Trejo, S.J., 1997, "Why Do Mexican Americans Earn Low Wages?" *Journal of Political Economy*, 105, pp. 1235-1268.

# 8 Data Appendix

As stated above, the primary data used for the analysis come from the NLSY79. The NLSY79 consists of 12,686 young men and women, living in the U.S., who were between the ages of 14 and 22 when the survey was first conducted in 1979. We focus our analysis on a set of Hispanics who were born in the U.S. or migrated to the U.S. before the age of seven. The NLSY79 respondents are asked their ethnicity in a series of six questions which identify their first (or only) through sixth ethnic identity. We focus on the first four questions which should account for paternal and maternal grandparents. We have constructed these ethnic variables such that they are mutually exclusive categories and we have identified an individual's Hispanic ethnicity as the first one indicated. We also created a variable to measure one's degree of "Hispanicity." An individual is classified as being "very" Hispanic if the only ethnicity he reports is Hispanic or if his first and second ethnic identities are Hispanic, "moderately" Hispanic if either of his first two ethnicities are Hispanic, and "less" Hispanic if only his third or fourth ethnicity indicated is Hispanic. The first category is intended to mainly include individuals for whom both parents are Hispanic, but also would include individuals whose only known parent is Hispanic. The second category is intended to include individuals for whom only one of the two known parents is Hispanic. The third category is intended to include those, not in the first two categories, for whom at least one grandparent is Hispanic. Respondents are asked if they were born in the U.S. or outside the U.S. For those individuals who were born abroad, there is a question eliciting year of entry into the U.S. This, coupled with the respondent's birth year, helped us identify individuals who moved here before the age of seven.

As stated earlier in the paper, the two outcome variables of interest in our analysis are the completed level of schooling and the ASVAB test scores. The final schooling level is constructed using the longitudinal data on highest grade completed, highest degree earned, enrollment status, and age. We followed Hansen et al. (2004) in the construction of this variable. Final schooling levels were constructed primarily using information on highest degree ever received in the most recent year such information was recorded. This question was asked beginning in 1988.[33] For individuals who were age 25 and above, if the highest degree ever received was: 1) an associate's degree, the individual was assigned 14 years of schooling; 2) a bachelor's degree, the individual was assigned 16 years of schooling; 3) a master's degree, the individual was assigned 18 years of schooling; and 4) a doctoral or professional degree, the individual was assigned 20 years of schooling. For individuals who indicated earning a high school diploma, but completed more than 12 years of schooling, we assigned them the years of completed schooling (provided it was less than 16 years). For individuals lacking degree information but who completed at least 12 years of school, we assigned them the highest grade completed. For individuals who were age 21 and above, if they indicated holding a high school diploma then they were assigned 12 years of schooling. For individuals who were missing degree information

[33]Note that the NLSY79 was an annual survey from 1979-1994 and from 1996–2000 the interviews were conducted biennially.

and completed less than 12 years of schooling, we assigned them the years of schooling completed. There were 36 people who remained; we were able to make reasonable judgements on 22 of these cases, and assigned them a number for the years of schooling completed.[34],[35]

The second, and primary, outcome variables of interest are the ASVAB test scores. The ASVAB test was administered to 11,914 (i.e. 94 percent) civilian and military NSLY79 respondents in 1980 and consists of 10 sections.[36] We standardize these test scores using the pooled Hispanic and White sample averages and standard deviations, and focus our attention on two math tests (arithmetic reasoning and math knowledge) and two verbal tests (word knowledge and paragraph comprehension) for reasons mentioned previously.

The other controls used in our analysis are as follows:[37] Family background measures include maternal and paternal schooling, the number of siblings, and whether an individual came from a "broken" home.[38] In order to maintain as large a sample as possible, we impute values for parental schooling when it is missing. We do so by regressing father's (mother's) schooling on his (her) spouse's schooling (to address the possibility of assortive mating), variables indicating being born abroad, and interactions with the ethnic indicators, along with some other controls. The predicted value is then imposed when information is missing on the father's (mother's) schooling level. A broken home is defined as one in which, at age 14, a respondent lived with someone other than his mother and father.[39] We also control for whether a respondent's mother or father was absent when he was age 14.[40] We control for parental country of birth with a variable indicating whether the birth was in a foreign country.[41] Our regional controls include a dummy variable for a southern residence and an urban residence at age 14. We also attempt to address any enclave-effects with the inclusion of the percent Hispanic living in an individual's county of residence at age 17. We use the figures provided in the 1980 Census in constructing this measure.

We control for the direct and opportunity costs of schooling with measures drawn from the 1980 Integrated Public Use Microdata Sample (IPUMS) and from the Department of Education's Integrated Postsecondary Education Data System (IPEDS) "Institutional Characteristics" 1980 survey. These variables

[34]For example, we encountered an individual who was age 21 in 1979 when the survey began. For years 1979-1987 he indicated having completed nine years of schooling. From 1988-1991, he reported 11 years. In 1992 he noted 13 years and from 1993-2000 he indicated 12 years. The information on highest degree completed was always missing. He also indicated not having a high school diploma or its equivalent for all years in the survey. Accordingly, we assigned this individual 12 years of schooling.

[35]Similarly for the Whites, we were able to reasonably eyeball 186 of 318 cases and assign a level of completed schooling.

[36]Many researchers proxy for ability in their regressions with the AFQT score which is a composite score derived from the tests listed above. Currie and Thomas (1999), however, argue that the AFQT score may be a better indicator of socioeconomic status than of intelligence. The AFQT is used by the Armed Forces and is designed to determine eligibility for enlistment and to assess an individual's trainability for service.

[37]Our control variables are similar to those employed by Hansen et al. (2004).

[38]While Hispanic parents may or may not choose to speak Spanish to their children, the children often speak English to each other. Of course the number of siblings can directly impact test scores in several ways such as through the resources invested in each child.

[39]A recent paper investigating the effect of divorce on cognitive and socioemotional development is Auginhaugh et al. (2005).

[40]Flouri and Buchanan (2004) find that parent involvement at age seven is an independent predictor of a child's educational attainment at age 20.

[41]Unlike for the children, we are unable to distinguish whether a parent was born in Puerto Rico or born in the rest of the U.S., and so any such births are considered domestic.

are collected for the county in which an individual lives when he is age 17, which was obtained from the NLSY79 Geocode files. Local labor market variables are constructed from the five percent sample of the IPUMS for prime-age (i.e. age 18-60) civilian, wage/salary employees. The IPUMS allows us to construct measures of the unemployment and wage rates by gender and schooling level.[42] Specifically, we consider the corresponding rates for individuals completing at least 12 years of school and for those who complete more than 12 years. The most detailed geographic identifier available in the IPUMS is a county group which is comprised of contiguous areas with a combined population of 100,000 or more residents; they may consist of actual county groups, but may also be single counties, cities, or Census-designated places. In order to construct unemployment rates that most reflected an individual's county of residence at age 17, we created a population-based weighted average of all the county groups in which an individual county was located. Our regressions include the difference between the average unemployment rate for individuals with more than 12 years of schooling and the average unemployment rate for individuals with 12 years of schooling or less. The wage rates are constructed by dividing the total income from wages and salary by the annual hours worked and averaging across individuals in a given county group. The annual hours worked are just the product of the weeks worked last year and the usual hours worked per week. These figures are again adjusted for the population of each county in the corresponding county group. Our analysis considers the difference between the average wage rate for individuals with more than 12 years of schooling and the average wage rate for individuals with 12 or less years of schooling. We used the IPEDS to obtain information on the location of all two- and four-year colleges, both public and private, in the U.S. in 1980.[43] Also included is the average in-state county and state tuition for public colleges. Specifically, we include the difference between the county and state tuition at each respective education level as in Cameron and Heckman (2001).

# 9 Technical Appendix

On all four tests the number of Hispanics who achieved the maximum score was sufficiently low that we did not correct for censoring. This was not the case for the non-Hispanics, however. Let $J_i$ if an individual's test score equals the maximum score on test $i$, $\bar{\tau}_i$. The observed test score of an individual, the equivalent of equation (22) in Section 3, now becomes:

$$\tau_i = \begin{cases} \theta_i \Delta + \beta_i X + \varepsilon_i \text{ if } I = 0, J = 1 \\ \theta_i \gamma Z + \lambda_i(\Delta - \gamma Z) + \beta_i X + (\theta_i - \lambda_i)u + \varepsilon_i \text{ if } I = 1, J = 1 \\ \bar{\tau}_i \text{ if } J = 0. \end{cases} \tag{32}$$

[42]The dollar figures are expressed in constant 1980 US$.

[43]The dummy variable corresponding to a two-year private college is omitted from the analysis because it is highly correlated with the variable indicating a two-year public college.

Assume that $P(J = 1) = P(\delta_i Z + \xi \leq 0) = \Phi(m_i)$, where $\xi \sim N(0, 1)$ and $m_i = -\delta_i Z$. The expected test score can be written as:

$$E(\tau_i) = \theta_i \left[\Delta\Phi(m_i) + (\gamma Z - \Delta)\Phi(c, m; \rho_{u\xi_i})\right] + \beta_i X[\Phi(m_i) + \lambda_i(\Delta - \gamma Z)\Phi(c, m_i; \rho_{u\xi i}) - \rho_{\varepsilon\xi_i}\phi(m_i)\sigma_{i\varepsilon} - \sigma_u(\theta_i - \lambda_i)\phi(c)\left((1 - \frac{\rho_{u\xi_i}\phi(m_i)}{\Phi(m_i)}\right) + \bar{\tau}_i(1 - \Phi(m_i)], \quad (33)$$

where $\Phi(c, m_i; \rho_{u\xi_i})$ is the bivariate normal distribution, $c = \frac{\Delta - \gamma Z}{\sigma_u}$ as before, $\rho_{\varepsilon\xi_i}$ is the correlation between $\varepsilon_i$ and $\xi_i$, and $\rho_{u\xi_i}$ is the correlation between $u$ and $\xi_i$. Estimation of $\Phi(c, m_i; \rho_{u\xi_i})$ for each combination of test score and schooling level would give rise to four distinct estimates of the parameters of $c$. To avoid this we carried out the following four-step procedure: First, we estimated each bivariate probit $\Phi(c, m_i; \rho_{u\xi_i})$ to obtain estimates of the correlation coefficient, $\rho_{u\xi_i}$. Next we then estimated the probits given by $\Phi(m_i)$ for each of the four tests and the probit given by $\Phi(c)$ for schooling. Third, we calculated the first-order Taylor series approximation to the probit given by $\Phi(c, m_i; \rho_{u\xi_i})$ as,

$$\Phi(c, m_i; \rho_{u\xi_i}) \approx \Phi(c)\Phi(m_i) - \frac{\rho_{u\xi_i}\phi(c)\phi(m_i)}{\Phi(c)\Phi(m_i)}. \quad (34)$$

We used our results from the first two steps to obtain estimates of $\rho_{u\xi_i}$, $\Phi(c)$, and $\Phi(m_i)$, and substituted these into equation (32) to obtain an estimate of $\Phi(c, m_i; \rho_{u\xi_i})$. And finally, we substituted all of the estimated values into equation (33) and estimated the equation by OLS.

TABLE 1
SELECT DESCRIPTIVE STATISTICS

| Variable | Hispanics (Nobs.=1,312) Mean | Std. Dev. | Hispanics who spoke Spanish at home as a child (Nobs.=1,124) Mean | Std. Dev. | Hispanics who did not speak Spanish at home as a child (Nobs.=188) Mean | Std. Dev. | Hispanics who parents were both born in the U.S. (Nobs.=624) Mean | Std. Dev. | Whites (Nobs.=2,940) Mean | Std. Dev. |
|---|---|---|---|---|---|---|---|---|---|---|
| Academic Achievement | | | | | | | | | | |
| **Schooling** | | | | | | | | | | |
| Completed schooling | 12.718 | 2.227 | 12.662 | 2.227 | 13.053 | 2.198 | 12.623 | 2.039 | 13.288 | 2.415 |
| Completed schooling at time of test | 11.232 | 1.769 | 11.184 | 1.749 | 11.521 | 1.860 | 11.298 | 1.675 | 11.797 | 1.780 |
| **Test Scores** | | | | | | | | | | |
| Arithmetic reasoning (max.=30) | 14.011 | 6.273 | 13.524 | 6.031 | 16.920 | 6.895 | 14.345 | 6.312 | 18.594 | 6.978 |
| Math knowledge (max.=25) | 10.621 | 5.650 | 10.268 | 5.450 | 12.734 | 6.342 | 10.601 | 5.614 | 14.027 | 6.229 |
| Word knowledge (max.=35) | 21.866 | 8.013 | 21.171 | 7.889 | 26.021 | 7.492 | 22.516 | 7.874 | 27.030 | 6.857 |
| Paragraph comprehension (max.=15) | 9.368 | 3.507 | 9.145 | 3.489 | 10.702 | 3.322 | 9.681 | 3.419 | 11.303 | 3.133 |
| Home Language Background | | | | | | | | | | |
| 1 if spoke Spanish at home as a child | 0.857 | 0.351 | 1.000 | 0.000 | 0.000 | 0.000 | 0.764 | 0.425 | 0.000 | 0.000 |
| Ethnicity | | | | | | | | | | |
| 1 if Cuban | 0.054 | 0.226 | 0.052 | 0.223 | 0.064 | 0.245 | 0.029 | 0.168 | 0.000 | 0.000 |
| 1 if Puerto Rican | 0.194 | 0.396 | 0.211 | 0.408 | 0.096 | 0.295 | 0.014 | 0.119 | 0.000 | 0.000 |
| 1 if other Hispanic | 0.136 | 0.343 | 0.099 | 0.298 | 0.356 | 0.480 | 0.170 | 0.376 | 0.000 | 0.000 |
| 1 if Mexican | 0.650 | 0.477 | 0.676 | 0.468 | 0.495 | 0.501 | 0.813 | 0.391 | 0.000 | 0.000 |
| Parental Schooling | | | | | | | | | | |
| Mother's schooling | 8.368 | 4.015 | 7.889 | 3.969 | 11.230 | 2.966 | 9.378 | 3.555 | 11.954 | 2.311 |
| Father's schooling | 8.659 | 4.413 | 8.166 | 4.352 | 11.609 | 3.550 | 9.694 | 3.913 | 12.247 | 3.302 |
| Average parents' schooling | 8.514 | 3.847 | 8.028 | 3.769 | 11.419 | 2.921 | 9.536 | 3.361 | 12.100 | 2.517 |
| 1 if average parents' schooling < 9 | 0.479 | 0.500 | 0.540 | 0.499 | 0.112 | 0.316 | 0.349 | 0.477 | 0.085 | 0.278 |
| 1 if 9 ≤ average parents' schooling < 12 | 0.301 | 0.459 | 0.286 | 0.452 | 0.388 | 0.489 | 0.385 | 0.487 | 0.292 | 0.455 |
| 1 if average parents' schooling ≥ 12 | 0.220 | 0.415 | 0.173 | 0.379 | 0.500 | 0.501 | 0.266 | 0.442 | 0.623 | 0.485 |

Note: For the full set of descriptive statistics, see Appendix Table A.1.

Source of data: *NLSY79, 1980 Census, 1980 IPUMS, 1980 IPEDS.*

TABLE 2

1st STAGE PROBIT: SELECT REGRESSION RESULTS FOR HISPANICS

| Dependent Variable: | Completed schooling | | |
|---|---|---|---|
| Estimation Technique: | Probit | | |
| | (1) | (2) | (3) |
| 1 if spoke Spanish at home as a child | -0.230 | 0.004 | --- |
| | (0.428) | (0.447) | |
| 1 if spoke Spanish at home as a child × 1 if average parents' schooling <9 | --- | --- | 0.0694 |
| | | | (0.676) |
| 1 if spoke Spanish at home as a child × 1 if 9 ≤ average parents' schooling < 12 | --- | --- | -0.221 |
| | | | (0.498) |
| 1 if spoke Spanish at home as a child × 1 if average parents' schooling ≥ 12 | --- | --- | 0.291 |
| | | | (0.572) |
| Mother's schooling | 0.038 | 0.036 | 0.033 |
| | (0.060) | (0.063) | (0.069) |
| Father's schooling | 0.164*** | 0.186*** | 0.186*** |
| | (0.049) | (0.047) | (0.057) |
| 1 if mother born abroad × mother's schooling | 0.019 | --- | --- |
| | (0.075) | | |
| 1 if father born abroad × father's schooling | -0.068 | --- | --- |
| | (0.071) | | |
| Pseudo $R^2$ | 0.383 | 0.430 | 0.432 |
| Log-likelihood | -554.220 | -240.100 | -239.112 |
| Nobs. | 1312 | 624 | 624 |

(Standard error)
*,**,***=significant at the 10, 5, and 1% level

Note: For full regression results, see Appendix Table B.1, columns 1, 2, and 3.

TABLE 3

$2^{nd}$ STAGE OLS: SELECT REGRESSION RESULTS FOR HISPANICS

EXOGENOUS CHOICE OF HOME LANGUAGE

| Dependent Variable: | Arithmetic reasoning | | | Math knowledge | | | Word knowledge | | | Paragraph comprehension | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Expected schooling at time of test | 0.118*** | 0.057 | 0.049 | 0.076** | 0.066 | 0.055 | 0.114*** | 0.062 | 0.066 | 0.107*** | 0.026 | 0.026 |
| | (0.030) | (0.040) | (0.040) | (0.030) | (0.042) | (0.042) | (0.031) | (0.044) | (0.043) | (0.033) | (0.046) | (0.045) |
| Expected potential experience at time of test | 0.025 | 0.042 | 0.042 | 0.012 | 0.014 | 0.017 | 0.086*** | 0.101*** | 0.099*** | 0.077*** | 0.097*** | 0.096*** |
| | (0.021) | (0.027) | (0.027) | (0.021) | (0.027) | (0.026) | (0.023) | (0.029) | (0.029) | (0.022) | (0.029) | (0.029) |
| 1 if spoke Spanish at home as a child | -0.255*** | -0.318*** | --- | -0.171** | -0.146 | --- | -0.264*** | -0.227** | --- | -0.124 | -0.115 | --- |
| | (0.069) | (0.092) | | (0.070) | (0.094) | | (0.080) | (0.097) | | (0.077) | (0.102) | |
| 1 if spoke Spanish at home as a child × 1 if average parents' schooling <9 | --- | --- | -0.032 | --- | --- | 0.087 | --- | --- | -0.244 | --- | --- | -0.059 |
| | | | (0.131) | | | (0.121) | | | (0.153) | | | (0.165) |
| 1 if spoke Spanish at home as a child × 1 if 9 ≤ average parents' schooling < 12 | --- | --- | -0.319*** | --- | --- | -0.155 | --- | --- | -0.270** | --- | --- | -0.108 |
| | | | (0.098) | | | (0.098) | | | (0.107) | | | (0.117) |
| 1 if spoke Spanish at home as a child × 1 if average parents' schooling ≥ 12 | --- | --- | -0.444*** | --- | --- | -0.233* | --- | --- | -0.162 | --- | --- | -0.152 |
| | | | (0.126) | | | (0.136) | | | (0.123) | | | (0.122) |
| Mother's schooling | 0.028*** | 0.031** | 0.047*** | 0.038*** | 0.037*** | 0.050*** | 0.058*** | 0.067*** | 0.065*** | 0.046*** | 0.053*** | 0.056*** |
| | (0.011) | (0.014) | (0.014) | (0.011) | (0.013) | (0.014) | (0.012) | (0.015) | (0.016) | (0.013) | (0.016) | (0.016) |
| Father's schooling | 0.046*** | 0.047*** | 0.066*** | 0.055*** | 0.058*** | 0.074*** | 0.057*** | 0.073*** | 0.071*** | 0.056*** | 0.069*** | 0.073*** |
| | (0.011) | (0.013) | (0.015) | (0.011) | (0.014) | (0.015) | (0.011) | (0.015) | (0.017) | (0.011) | (0.014) | (0.016) |
| 1 if mother born abroad × mother's schooling | -0.004 | --- | --- | -0.006 | --- | --- | -0.023* | --- | --- | -0.022 | --- | --- |
| | (0.012) | | | (0.013) | | | (0.014) | | | (0.014) | | |
| 1 if father born abroad × father's schooling | -0.031*** | --- | --- | -0.033*** | --- | --- | -0.033*** | --- | --- | -0.025** | --- | --- |
| | (0.011) | | | (0.012) | | | (0.013) | | | (0.013) | | |
| $R^2$ | 0.200 | 0.219 | 0.230 | 0.188 | 0.226 | 0.233 | 0.259 | 0.226 | 0.303 | 0.199 | 0.216 | 0.216 |
| Nobs. | 1312 | 624 | 624 | 1312 | 624 | 624 | 1312 | 624 | 624 | 1312 | 624 | 624 |

(Bootstrapped standard error)

*,**,***=significant at the 10, 5, and 1% level

Notes: Test scores are standardized. For full regression results, see Appendix Table B.2.

Source of data: *NLSY79, 1980 Census, 1980 IPUMS, 1980 IPEDS*.

TABLE 4
BLINDER-OAXACA DECOMPOSITIONS FOR STANDARDIZED TEST SCORES
EXOGENOUS CHOICE OF LANGUAGE

BASED ON REGRESSIONS FOR ENTIRE HISPANIC SAMPLE WHERE EFFECT OF SPEAKING SPANISH AT HOME IS NOT ALLOWED TO VARY WITH PARENTAL SCHOOLING

| | Arithmetic reasoning | Math knowledge | Word knowledge | Paragraph comprehension |
|---|---|---|---|---|
| | (1) | (4) | (7) | (10) |
| Covariates | | | | |
| Parental schooling | 0.328 | 0.350 | 0.384 | 0.320 |
| Other covariates | 0.244 | 0.329 | 0.249 | 0.260 |
| Total covariates | 0.573 | 0.679 | 0.633 | 0.580 |
| Parameters | | | | |
| Speaking Spanish at home | 0.218 | 0.147 | 0.226 | 0.106 |
| Parental foreign birth | -0.148 | -0.248 | -0.280 | -0.193 |
| Other parameters | 0.004 | -0.034 | 0.100 | 0.082 |
| Total parameters | 0.074 | -0.135 | 0.046 | -0.006 |
| Total | 0.647 | 0.545 | 0.679 | 0.574 |

BASED ON REGRESSIONS FOR HISPANIC SUBSAMPLE WHERE EFFECT OF SPEAKING SPANISH AT HOME IS ALLOWED TO VARY WITH PARENTAL SCHOOLING

| | Arithmetic reasoning | Math knowledge | Word knowledge | Paragraph comprehension |
|---|---|---|---|---|
| | (2) | (5) | (8) | (11) |
| Covariates | | | | |
| Parental schooling | 0.218 | 0.231 | 0.225 | 0.181 |
| Other covariates | 0.238 | 0.316 | 0.225 | 0.232 |
| Total covariates | 0.456 | 0.547 | 0.450 | 0.413 |
| Parameters | | | | |
| Speaking Spanish at home | 0.243 | 0.111 | 0.173 | 0.088 |
| Parental foreign birth | 0.000 | 0.000 | 0.000 | 0.000 |
| Other paramters | -0.099 | -0.110 | -0.030 | -0.019 |
| Total parameters | 0.143 | 0.001 | 0.143 | 0.068 |
| Total | 0.600 | 0.548 | 0.593 | 0.481 |

BASED ON REGRESSIONS FOR HISPANIC SUBSAMPLE WHERE EFFECT OF SPEAKING SPANISH AT HOME IS ALLOWED TO VARY WITH PARENTAL SCHOOLING

| | Arithmetic reasoning | Math knowledge | Word knowledge | Paragraph comprehension |
|---|---|---|---|---|
| | (3) | (6) | (9) | (12) |
| Covariates | | | | |
| Parental schooling | 0.218 | 0.231 | 0.225 | 0.181 |
| Other covariates | 0.238 | 0.316 | 0.225 | 0.232 |
| Total covariates | 0.456 | 0.547 | 0.450 | 0.413 |
| Parameters | | | | |
| Speaking Spanish at home | 0.168 | 0.051 | 0.182 | 0.073 |
| Parental foreign birth | 0.000 | 0.000 | 0.000 | 0.000 |
| Other parameters | -0.025 | -0.050 | -0.039 | -0.005 |
| Total parameters | 0.143 | 0.001 | 0.143 | 0.068 |
| Total | 0.600 | 0.548 | 0.593 | 0.481 |

Note: For full regression results, see Appendix Tables B.2 and B.5.

Source of data: *NLSY79, 1980 Census, 1980 IPUMS, 1980 IPEDS*.

TABLE 5

$2^{nd}$ STAGE OLS: SELECT REGRESSION RESULTS FOR HISPANICS

ENDOGENOUS CHOICE OF HOME LANGUAGE

| Dependent Variable: | Arithmetic reasoning | | | Math knowledge | | | Word knowledge | | | Paragraph comprehension | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Expected schooling at time of test | 0.118*** | 0.062 | 0.056 | 0.080*** | 0.073* | 0.070* | 0.114*** | 0.062 | 0.064 | 0.097*** | 0.022 | 0.020 |
| | (0.029) | (0.039) | (0.039) | (0.030) | (0.039) | (0.039) | (0.031) | (0.043) | (0.043) | (0.032) | (0.044) | (0.044) |
| Expected potential experience at time of test | 0.021 | 0.032 | 0.033 | 0.007 | 0.008 | 0.007 | 0.082*** | 0.097*** | 0.096*** | 0.081*** | 0.096*** | 0.096*** |
| | (0.020) | (0.026) | (0.026) | (0.021) | (0.026) | (0.026) | (0.022) | (0.029) | (0.029) | (0.022) | (0.028) | (0.028) |
| 1 if spoke Spanish at home as a child | -0.731*** | -0.882*** | --- | -0.379** | -0.274 | --- | -0.518*** | -0.516** | --- | -0.413** | -0.434* | --- |
| | (0.165) | (0.214) | | (0.176) | (0.221) | | (0.173) | (0.238) | | (0.180) | (0.231) | |
| 1 if spoke Spanish at home as a child × 1 if average parents' schooling <9 | --- | --- | -0.591** | --- | --- | -0.069 | --- | --- | -0.511* | --- | --- | -0.380 |
| | | | (0.250) | | | (0.241) | | | (0.291) | | | (0.288) |
| 1 if spoke Spanish at home as a child × 1 if 9 ≤ average parents' schooling < 12 | --- | --- | -0.877*** | --- | --- | -0.307 | --- | --- | -0.543** | --- | --- | -0.430* |
| | | | (0.216) | | | (0.222) | | | (0.243) | | | (0.240) |
| 1 if spoke Spanish at home as a child × 1 if average parents' schooling ≥ 12 | --- | --- | -0.947*** | --- | --- | -0.258 | --- | --- | -0.471* | --- | --- | -0.451* |
| | | | (0.240) | | | (0.250) | | | (0.257) | | | (0.244) |
| Mother's schooling | 0.018 | 0.019 | 0.035** | 0.033*** | 0.034** | 0.043*** | 0.052*** | 0.061*** | 0.060*** | 0.040*** | 0.047*** | 0.050*** |
| | (0.011) | (0.015) | (0.015) | (0.011) | (0.014) | (0.014) | (0.013) | (0.016) | (0.018) | (0.013) | (0.017) | (0.017) |
| Father's schooling | 0.039*** | 0.040*** | 0.058*** | 0.051*** | 0.056*** | 0.067*** | 0.053*** | 0.070*** | 0.069*** | 0.054*** | 0.066*** | 0.070*** |
| | (0.011) | (0.014) | (0.016) | (0.012) | (0.014) | (0.015) | (0.012) | (0.016) | (0.017) | (0.011) | (0.014) | (0.018) |
| 1 if mother born abroad × mother's schooling | 0.006 | --- | --- | -0.001 | --- | --- | -0.018 | --- | --- | -0.016 | --- | --- |
| | (0.013) | | | (0.013) | | | (0.015) | | | (0.015) | | |
| 1 if father born abroad × father's schooling | -0.024** | --- | --- | -0.030** | --- | --- | -0.029** | --- | --- | -0.021* | --- | --- |
| | (0.012) | | | (0.012) | | | (0.013) | | | (0.013) | | |
| $R^2$ | 0.204 | 0.226 | 0.233 | 0.188 | 0.226 | 0.230 | 0.257 | 0.302 | 0.302 | 0.201 | 0.219 | 0.219 |
| Nobs. | 1312 | 624 | 624 | 1312 | 624 | 624 | 1312 | 624 | 624 | 1312 | 624 | 624 |

(Bootstrapped standard error)

*,**,***=significant at the 10, 5, and 1% level

Notes: Test scores are standardized. For full regression results, see Appendix Table B.4.

Source of data: *NLSY79, 1980 Census, 1980 IPUMS, 1980 IPEDS.*

TABLE 6
BLINDER-OAXACA DECOMPOSITIONS FOR STANDARDIZED TEST SCORES
ENDOGENOUS CHOICE OF LANGUAGE

BASED ON REGRESSIONS FOR ENTIRE HISPANIC SAMPLE WHERE EFFECT OF SPEAKING SPANISH AT HOME IS NOT ALLOWED TO VARY WITH PARENTAL SCHOOLING

| | Arithmetic reasoning | Math knowledge | Word knowledge | Paragraph comprehension |
|---|---|---|---|---|
| | (1) | (4) | (7) | (10) |
| Covariates | | | | |
| Parental schooling | 0.328 | 0.350 | 0.384 | 0.320 |
| Other covariates | 0.244 | 0.329 | 0.249 | 0.260 |
| Total covariates | 0.573 | 0.679 | 0.633 | 0.580 |
| Parameters | | | | |
| Speaking Spanish at home | 0.626 | 0.324 | 0.443 | 0.354 |
| Parental foreign birth | -0.170 | -0.257 | -0.291 | -0.207 |
| Other paramters | -0.382 | -0.201 | -0.106 | -0.152 |
| Total parameters | 0.074 | -0.135 | 0.046 | -0.006 |
| Total | 0.647 | 0.545 | 0.679 | 0.574 |

BASED ON REGRESSIONS FOR HISPANIC SUBSAMPLE WHERE EFFECT OF SPEAKING SPANISH AT HOME IS ALLOWED TO VARY WITH PARENTAL SCHOOLING

| | Arithmetic reasoning | Math knowledge | Word knowledge | Paragraph comprehension |
|---|---|---|---|---|
| | (2) | (5) | (8) | (11) |
| Covariates | | | | |
| Parental schooling | 0.218 | 0.231 | 0.225 | 0.181 |
| Other covariates | 0.238 | 0.316 | 0.224 | 0.231 |
| Total covariates | 0.456 | 0.547 | 0.450 | 0.413 |
| Parameters | | | | |
| Speaking Spanish at home | 0.672 | 0.209 | 0.393 | 0.331 |
| Parental foreign birth | 0.000 | 0.000 | 0.000 | 0.000 |
| Other paramters | -0.529 | -0.208 | -0.249 | -0.262 |
| Total parameters | 0.144 | 0.001 | 0.143 | 0.069 |
| Total | 0.600 | 0.548 | 0.593 | 0.481 |

BASED ON REGRESSIONS FOR HISPANIC SUBSAMPLE WHERE EFFECT OF SPEAKING SPANISH AT HOME IS NOT ALLOWED TO VARY WITH PARENTAL SCHOOLING

| | Arithmetic reasoning | Math knowledge | Word knowledge | Paragraph comprehension |
|---|---|---|---|---|
| | (3) | (6) | (9) | (12) |
| Covariates | | | | |
| Parental schooling | 0.218 | 0.231 | 0.225 | 0.181 |
| Other covariates | 0.238 | 0.316 | 0.224 | 0.231 |
| Total covariates | 0.456 | 0.547 | 0.450 | 0.413 |
| Parameters | | | | |
| Speaking Spanish at home | 0.587 | 0.150 | 0.392 | 0.315 |
| Parental foreign birth | 0.000 | 0.000 | 0.000 | 0.000 |
| Other paramters | -0.444 | -0.149 | -0.249 | -0.246 |
| Total parameters | 0.144 | 0.001 | 0.143 | 0.069 |
| Total | 0.600 | 0.548 | 0.593 | 0.481 |

Note: For full regression results, see Appendix Tables B.3 and B.5.

Source of data: *NLSY79, 1980 Census, 1980 IPUMS, 1980 IPEDS.*