WORKING PAPERS
RESEARCH DEPARTMENT

WORKING PAPER NO. 02-16
THE RETURNS TO SPEAKING A SECOND LANGUAGE

Albert Saiz
Federal Reserve Bank of Philadelphia
Harvard University

Elena Zoido
Harvard University

October 2002

FEDERAL RESERVE BANK OF PHILADELPHIA
Ten Independence Mall, Philadelphia, PA 19106-1574• (215) 574-6428• www.phil.frb.org

WORKING PAPER NO. 02-16

THE RETURNS TO SPEAKING A SECOND LANGUAGE*

Albert Saiz†
Federal Reserve Bank of Philadelphia and Harvard University

Elena Zoido‡
Harvard University

October 2002

* The views expressed in this paper are those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. We want to thank Larry Katz for guidance. David Cutler, Caroline Hoxby, John McHale, Uri Simonsohn, and participants at the Harvard Labor/Public Finance Workshop provided very helpful comments. Bridget Terry Long generously shared with us college quality data.

† E-mail: Albert.Saiz@phil.frb.org
‡ E-mail: ezoido@post.harvard.edu

Abstract

Does speaking a foreign language have an impact on earnings? We use a variety of empirical strategies to address this issue for a representative sample of U.S. college graduates. OLS regressions with a complete set of controls to minimize concerns about omitted variable biases, propensity score methods, and panel data techniques all lead to similar conclusions. The hourly earnings of those who speak a foreign language are more than 2 percent higher than the earnings of those who do not. We obtain higher and more imprecise point estimates using state high school graduation and college entry and graduation requirements as instrumental variables.

JEL: J31, I29.
Keywords: Second language, earnings, human capital.

1 Introduction

Does learning a second language in the United States pay off? This is a question that most Americans face and answer in their education years. It is more relevant as the use of English as *the* international common language becomes more extended. Learning a second language is a possibility many consider at some point. Language courses are an option or a requirement in elementary, secondary, and college education. Private foreign language institutions, educational materials, and distance courses are regularly advertised in newspapers and magazines. Reports in the popular press suggest that speaking a second language has become a "desirable skill in the work force, one employers are willing to pay for."[1] Despite the salience of the decision of whether to invest in a second language and the anecdotal evidence that suggests that it is a valuable skill in the labor market, little research has been undertaken on the labor market returns to learning a second language in the United States. The literature on language acquisition and labor market outcomes has so far concentrated on the returns to learning English for immigrants.[2] In this paper, we estimate the returns to speaking a second language for college graduates who are native English speakers in the U.S.

We find that college graduates who speak a second language earn, on average, wages that are 2 percent higher than those who don't. We include a complete set of controls for general ability using information on grades and college admission tests and reduce the concern that selection drives the results controlling for the academic major chosen by the student. We obtain similar results with simple regression methods if we use nonparametric methods based on the propensity score and if we exploit the temporal variation in the knowledge of a second language. The estimates, thus, are not driven by observable differences in the composition of the pools of bilinguals and monolinguals, by the linear functional form that we impose in OLS regressions, or by constant unobserved heterogeneity. To reduce the concern that omitted variables bias our estimates, we make use of several instrumental variables (IVs). Using high school and college graduation requirements as instruments, we estimate more substantial returns to learning a second language, on the order of 14 to 30 percent. These results have high standard errors, but they suggest that OLS estimates may actually be biased downward.

The estimation of the returns to speaking foreign languages may have important policy impli-

[1] *New York Times*, October 18, 1999.

[2] Some examples are Carliner (1995) and Chiswick and Miller (1997b, 1998). Section 3 discusses this literature in more detail.

cations. Many states regularly update their high school graduation requirements. There is much discussion on the skills that secondary education should provide. Skills rewarded in the labor market seem of paramount importance among these. The relative labor market value of second language skills should thus weigh heavily in these choices. Colleges and other educational institutions also need to make decisions on graduation requirements. This paper should provide helpful input to these issues.

Furthermore, our results may help us understand individual decisions on whether to learn a second language. Human capital theory guides our understanding of the individual's decision. One should invest in the acquisition of a foreign language if the present value of the future returns for doing so exceeds the costs. Some of the returns from learning a second language consist of the direct consumption of services produced by the individual's knowledge of the language. Speaking a second language while travelling abroad, asking an immigrant shopkeeper for a product in her native language, and relating to foreign friends are all examples of these. This paper cannot address them. The value of the labor market skills that learning a foreign language provides is the other important component of the returns to learning the language. Our objective is to quantify it.

There are several reasons that can make the knowledge of a foreign language useful in the American labor market. First, American companies export and import products to the rest of the world. Knowledge of a second language may be a valuable asset for such companies doing business abroad or for companies catering to immigrants or people of foreign background within the U.S. Offering a service in the language of the prospective customers may provide a competitive advantage for firms, especially in areas where the foreign born tend to settle. Finally, second language skills are in demand from several government agencies that deal with foreign affairs: diplomatic service, CIA, military intelligence, and the like.

Additionally, an extensive literature suggests that learning a second language may help individuals develop their cognitive and communicative abilities. Research in linguistics underlines the possible advantages of bilingualism in terms of intellectual and academic achievement.[3] If speaking a second language is important for improving cognitive capabilities, we should find that the individuals who speak a second language are more productive and earn higher wages. At the same

[3] Cooper (1987) finds that math and verbal scores are higher with each additional year of foreign language study. Olsen and Brown (1992) show that "students who had completed a foreign language course in high school tended to have higher scores on the ACT exams in English and math *regardless of their ability level.*" Learning a foreign language may help develop analytic and interpretative capacities. Cook (1997) reports that "increased metalinguistic awareness of phonology, syntax, and the arbitrary nature of meaning, and gains in cognitive flexibility" are established outcomes from learning a second language. This literature, though, faces challenges similar to the ones we face in our research. Section 3 presents a detailed discussion of these issues.

time, access to foreign media and literature may help innovation and adoption of best practices from abroad and improve workers' productivity.

The paper proceeds as follows. In Section 2 we discuss the existing literature on the returns to learning a second language. Section 3 addresses the potential methodological problems that estimating the returns to a second language faces. In Section 4, we introduce the data sources and describe the characteristics of individuals speaking a second language in our sample. Section 5 presents the results of the OLS, propensity scores, panel data, and IV approaches. Section 6 concludes.

2 The Literature on Second Language and Earnings

Economic analysis has been applied only recently to our understanding of the individual decision to invest in a second language. Grenier and Vaillancourt (1983) were the first to identify foreign languages as an element in the human capital portfolio of individuals. They use the framework laid out by Becker (1964) and Mincer (1974) to describe how individual characteristics have an impact on the gross costs, the gross returns and the information on the value of the returns to this investment. Similarly, Ridler and Pons-Ridler (1984) analyze the decision to learn a second language as an investment decision. These authors emphasize the importance of second languages as consumption goods and discuss the economic costs of forced "francisation" in Quebec as an economic problem.

So far, the empirical literature on the returns to speaking a second language has concentrated on the incentives and wage premia that immigrants receive for mastering the language of their country of adoption. Carliner (1995) finds that most immigrants in the U.S. are proficient in English. He shows that each additional year of residence in the U.S. increases the probability of being proficient in English by 1.1 percentage points. Chiswick and Miller (1997a, 1998) show that the probability of speaking English for immigrants in the U.S. responds to the costs and benefits of doing so. Chiswick and Miller (1997b) study the labor market returns to English acquisition among immigrants in the United States. They find that foreign-born non-English speaking immigrants who are fluent in English earn about 15 to 19 percent more than the ones who are not. Similar wage premia are observed in other countries. Chiswick (1997), for example, analyzes the wage differential for immigrants who speak Hebrew in Israel. In Canada, Chiswick and Miller (2000) find that immigrants who can speak one of the official languages earn wages that are 10 to 12 percent higher. They find evidence of positive selection: individuals with higher tendency to learn an official language tend to earn higher wages, possibly for reasons other than such proficiency.

Despite the extensive literature on the cognitive and developmental effects of bilingualism, there is little research on the returns to speaking a second language for nonimmigrants.[4,5] An important and motivating finding for us is the fact that foreign language courses in high school do affect wages (Altonji, 1995). Indeed, his results show that foreign language courses have higher returns than courses in mathematics, science, and verbal skills. Such strong effects of taking second language courses on future wages seem puzzling in the American context. These results could be driven by the correlation between taking second language courses and omitted variables, such as high school quality or family background, or by selection. One would expect, for example, that foreign language courses are more sensitive to the affluence of the families whose children attend a school than math, science, or language courses are. A second possibility, one that we are interested in exploring, is that speaking a second language is a valuable skill in the labor market.

Finally, and closest to the spirit of our work, Lopez (1999) estimates the labor market returns to speaking a second language in the U.S. He uses observations form the National Adult Literacy survey (NALS) of 1992. The main shortcoming of this data set is that only individuals belonging to a language minority were questioned regarding their language skills. Native English speakers who have subsequently learned a foreign language, the treatment group of interest for us, are excluded from the questionnaire on second languages. Thus, this author's basic test is whether language *retention* among language minorities is associated with higher or lower labor market earnings. He finds that those who speak a second language proficiently earn wages that are 14.2 percent higher than those earned by people in language minorities who do not speak the minority language well. He also finds that minority language individuals who speak their mother tongue well earn 13.5 percent more than English monolinguals, but the difference is not significant.

[4] See Cook (1997) for a survey on the benefits of bilingualism on the development of cognitive ability and communication skills.

[5] In the Canadian context, Shapiro and Stelcner (1997) and Christofides and Swidinsky (1998) explore the returns to speaking a second official language. Pendakur and Pendakur (1998) estimate the returns of speaking French in Vancouver and Toronto and of speaking English in Francophone Montreal. These authors find that speaking a nonofficial language is associated with a *negative* earnings differential. The interpretation of these results is subject to several problems. Selection into learning a language different from the official ones may affect these estimates dramatically. For example, if immigrants speak other languages in much greater proportions than the general population, it will be difficult to disentangle the effects of assimilation from the pure effects of speaking more languages. This paper points to the fact that selection into learning a second language can be correlated with individual attributes that decrease earnings, which suggests the importance of controlling for potentially confounding factors.

3 Empirical Strategy

Measuring the returns to learning a foreign language presents considerable methodological challenges, in common with the work on the returns to schooling. Think of a running a regression like the following:

$$\log w_i = \alpha + \beta X_i + \gamma SL_i + \epsilon_i$$

where $\log w$ is the log of earnings, X is a vector of personal characteristics, and SL is an indicator of whether the individual speaks a second language.

Obviously, speaking a second language is not randomly assigned in the population. Individuals choose whether to learn a foreign language according to their potential earnings. Thus, a selection problem arises. Consider the distribution of the benefits of speaking a second language among the population. Assume that the costs of learning it are the same for all individuals. Then, those individuals with higher returns to speaking a second language will have greater incentives to invest in learning it. Empirically, we may find that the individuals who speak a second language obtain higher earnings. The average returns of speaking it, if the second language skills were randomly assigned, would be smaller. Selection by earnings will, thus, bias conventional estimates upward. This problem is akin to the selection problem in the labor market participation decision of female workers (Heckman, 1980).

Additionally, the costs of learning a second language may be smaller for individuals with higher cognitive competence. If this is the case, more able individuals will learn foreign languages with a higher probability. If we do not correct for this differential selection with respect to ability, our estimates will contain an ability bias, as the coefficient on the foreign language variable will capture part of the effect of ability on earnings (Griliches, 1977). Again, this causes the estimates of the returns to a second language to be biased upward.

We also need to take into account that there may be other unobservable characteristics of the individuals that simultaneously determine earnings and whether an individual speaks a second language. People who study foreign languages may have different preferences regarding the types of jobs they like and the kind of studies they choose. Consider the population of college graduates. People in humanities majors study foreign languages with higher likelihood and are more likely to choose teaching jobs, for example. This selection problem would again bias our estimates, although the direction of the bias is not clear *a priori*.

An additional problem of the OLS specification is that it imposes a linear relation among the variables. The returns to learning a second language, however, may vary markedly among different groups. If the effect of the treatment we are trying to estimate (the effects for those who learn a foreign language of doing so) is contingent on observable individual characteristics, simple OLS regressions may not yield unbiased estimates. We can think of this as a problem of model specification. The coefficient of the foreign language indicator cannot be interpreted independently of the other characteristics of the individual. A fully interacted model may not be the solution because we do not know ex-ante the relevant interactions and the functional form of the treatment effect. Moreover, if the treatment group (the people who speak a second language) and the control group (those who do not) are different in their observable characteristics, the coefficient of a treatment indicator may be very sensitive to the specification (Lalonde, 1986).

We address all these concerns through several complementary approaches in our empirical implementation. First, we include a very complete set of control variables in our OLS regressions to reduce the concern that selection may bias our estimates. We try to mitigate the ability biases by including SAT scores, GPA, parental education, and indicators of the quality of the college attended in our regressions. We include controls for the major chosen by the individuals to capture unobserved individuals' characteristics, as the choice of a college major may be viewed as an indicator of career preferences. Second, we address the possibility that the linearity imposed in the OLS regression biases our estimates of the average effect of speaking a second language on earnings. We examine the sensitivity of the results to this assumption by using several methods based on the propensity score: weighting, stratification, and matching.

We will be able to give a causal interpretation to the estimates obtained by OLS regression and propensity score methods only if, once we condition on the observable characteristics of the individuals in our sample, speaking a foreign language or not is independent of an individual's potential earnings. This may not be the case. To address the problem of constant unobserved individual characteristics that may bias our results, we exploit the longitudinal dimension of the data set. We thus compare the evolution of earnings for people who learned a second language between 1993 and 1997 to the evolution of earnings for other groups. Finally, we address the "selection by earnings" problem (and the rest of omitted variables that we may have not succeeded in controlling for) using instrumental variables (IV). High school foreign language requirements in the state where the respondent attended high school and college second language requirements are the instruments used.

4 Data

We focus on the returns to speaking a second language in the U.S. There are no available data sources that contain information on foreign language proficiency for a representative sample of the American population. Our main data source is the "Baccalaureate and Beyond Longitudinal Study" (B&B henceforth), from the National Center for Education Statistics, which contains unique information on second language ability. This data set tracks the experiences of a nationally representative cohort of college graduates who received their bachelor's degrees during the 1992-1993 academic year. It contains information on the characteristics of the universities the individuals attended (four-year institutions in all cases), from the students' undergraduate transcripts, and from answers to several questions posed to the individuals in surveys carried out in 1993, 1994, and 1997. These questions include information on demographics, parents' background, earnings and job characteristics, academic major, and courses taken in college and after graduation. The sample included 11,192 students initially. Of these, 9,274 individuals responded to the initial interview and all follow ups.

Respondents were asked in 1993 and 1997 whether they spoke a foreign language and to identify which language. The question used was, "Do you have conversational knowledge of languages other than English?" If the answer was affirmative, the interviewer followed up with the question "What are these languages?" This question is ideal for us, as it is intended to capture knowledge of a second language and not whether an individual belongs to a language minority. This feature contrasts with other data sets, which contain questions on second language only for linguistic minorities.

Our choice of this data set is therefore very much driven by the existence of a second language question geared toward native speakers. Other features of B&B, however, make it very suitable for this paper because it includes variables that can be considered cognitive ability proxies, namely SAT scores and college grade point average (GPA). It provides information on where the parents of the individual were born, whether the respondent spoke English at home while growing up, and whether she is an American citizen. We will use these variables to focus our analysis on individuals born to parents born in the U.S., American citizens who spoke English as their mother tongue. B&B also contains a wealth of information on academic and job careers and on earnings. The only shortcoming of our using this data source is that the results we obtain are representative only for the population of college graduates. Thus, we suggest caution with their generalization. It is not clear, a priori, if one should expect greater or smaller effects on the general population.

We complement B&B with some additional information. We use an indicator of college quality compiled by Caroline Hoxby and Bridget Terry Long.[6] This is a classification of colleges into six categories according to the average SAT of admitted students in the college and the average high school GPA of attending students. Per capita income in the states of residence in 1989, 1993, and 1997 is from the Bureau of Labor Statistics. The percentage of Hispanic residents in a State are obtained for 1989, 1993, and 1997 from the U.S. Census Population Estimates, available at www.census.gov.

In our IV estimation we exploit variation in high school and college graduation requirements. High school graduation requirements in 1989 (the estimated year for which our college seniors graduated from high school) are obtained from the *Digest of Education Statistics*. We gathered information on second language college requirements from the Modern Language Association (MLA) Survey of Foreign Language Entrance and Degree Requirements (FLEDR) (Brod and Lapointe, 1989). The FLEDR was undertaken in July 1988. Questionnaires were mailed to institutions of higher education that offered courses in one foreign language or more. The response rate was 98.3 percent. The survey contains data on whether an institution has foreign language admission and graduation requirements and which majors and degrees (e.g., BA or BS) require foreign languages for graduation. The survey also includes the number of semesters or semester hours of a second language required for graduation, if any. We concentrate on graduation requirements and focus on whether an institution has any foreign language requirement. This avoids endogeneity problems, as opposed to using an indicator on whether the major and degree of the student in the sample required a foreign language: people self-select into majors and degrees. We use these variables as instruments that should help predict whether a college graduate has knowledge of a second language.

The basic demographic characteristics of the individuals in our sample are presented in Table A.1. in the Appendix. The table is based on the 1997 wave of the survey. Panel A shows that the average individual is around 30 years old in this second survey and that she has more than 3 and a half years of labor market experience. A third of the individuals are married; less than half of them are men. More than a quarter of the individuals have received some graduate degree, and most of them (67 percent) attended a public college to obtain their degree. Thirty-four percent of the individuals in the sample claim to speak a foreign language or more.[7] Panel B shows the regional

[6] See Appendix A and Hoxby and Long (1999).

[7] The proportion is slightly lower (28 percent) in the subsample of individuals who spoke English at home while growing up, whose parents were born in the US and who are American citizens.

distribution of the individuals in the sample.

4.1 Who Speaks a Foreign Language?

We start our analysis with a brief description of the characteristics of the individuals who speak one or more foreign languages. Table 1 compares individuals who speak and who do not speak foreign languages. Panel A reveals some differences between these two groups: the probability of working, having attended a public college, holding an MBA degree, and being married is higher among those who speak English only. They are older and have slightly more work experience.[8] On the other hand, those who speak a foreign language are more likely to hold other graduate degrees, are disproportionately more likely to be black, and have slightly higher wages.[9] Interestingly, those who speak a language other than English are relatively more likely to have high SAT scores (Panel B). This suggests that those who decide to learn a foreign language may have higher cognitive skills.

The proportion of residents who speak a foreign language varies by region, as Panel C in Table 1 shows. New England and the Pacific region have the highest proportion of residents who speak a second language (41 percent). The percentage of foreign language speakers is lowest in East North Central (IA, KS, MN, MO, NE, ND, SD) and East South Central (AL, KY, MS, TN).

Note that more than 15 percent of those who speak a foreign language actually speak more than one language other than English (Table 2, Panel A). Around 14 percent speak two foreign languages, almost 2 percent speak three. We find no differences by gender in the number of languages spoken. In Panel B, we observe that most of them speak Spanish (58 percent), followed by French and German. Four percent of the individuals in our sample speak an eastern or southeastern Asian language, and 2 percent speak Chinese.[10] Spanish remains the most popular language in all regions and for both men and women. In relative terms, a higher percentage of women choose French and Italian as their second language, while men prefer German (the differences are statistically significant). Different languages are spoken by residents of different regions: Spanish is relatively more popular as a foreign language in the West South Central, East South Central, and Mountain

[8] Experience is calculated by, first, obtaining the months between graduation and the interview date. We then subtract the months of full-time graduate study that the interviewee reports.

[9] We obtain hourly wages, the dependent variable of interest to us, by dividing yearly, monthly, weekly, or daily wages by the total number of hours the individual reports to work in the relevant period. We find some extreme values in earnings and adjust the sample by truncating the sample at the 1 percent. That is, we drop from the sample the observations in either tail of the distribution. See Angrist and Krueger (1999) for an overview of these trimming methods as a way to reduce the biases due to measurement error.

[10] The proportions are very similar in the subsample of native English speakers with parents born in the U.S. 63 percent speak Spanish, 24 percent choose French, 12 percent German. Only 0.6 percent speak an Eastern or Southeastern Asian language, and 0.4 percent speak Chinese.

regions; residents in the South Atlantic region and New England speak French more frequently, while German speakers are disproportionately located in the West North Central region.

5 The Returns to Speaking a Second Language

5.1 OLS Results

We begin our investigation of the returns to speaking a second language with a variety of conventional earnings functions estimated by OLS. Table 3 reports the estimates. We include quadratic functions of age and experience, and race, gender, marital status and parental education controls. We add controls for the income per capita in the state of residence[11], for college quality, for whether the individual holds a graduate degree (MBA, Ph.D., and other master's) and ability proxies (normalized GPA and SAT quartile)[12]. It is particularly important to account for the major in college. In appendix Table A.2 we illustrate how graduates from majors in which the percentage of individuals who report speaking a second language is higher are also the majors in which the average wages are lower.

The estimated foreign language coefficient implies a 2.8 percent wage premium associated with speaking a foreign language for the average individual in the sample (column 1). We estimate additional models to test the robustness of this estimate. In particular, we restrict our sample to those individuals whose native language is English and account for the possibility that some of the individuals in our sample might be part-time workers. In the regressions shown in the next two columns, we restrict the sample to individuals who spoke English at home while growing up (column 2). Column 3 focuses on individuals who spoke English at home while growing up, are American citizens, and whose parents were born in the U.S. (more than 80 percent of the sample). Finally, column 4 shows the returns for full-time workers who spoke English at home when they were children.[13] Again, the point estimates are similar, although slightly lower: 1.9-2.2 percent. These compare with estimates in the range of 0.9 percent when we exclude the controls for the college major chosen by the individual.

In separate (unreported) regressions, we explore the labor market returns to speaking specific languages. We estimate OLS regressions following the previous specifications but allow the coefficient to vary by language spoken. In our sample, German is the language that obtains the highest

[11] This yields the same results as the addition of state of residence dummies.

[12] Results do not change if we control for industry and occupation dummies.

[13] We define an individual as a full-time worker if she works more than 35 hours a week.

rewards in the labor market. The returns to speaking German are 3.8 percent, while they are 2.3 for speaking French and 1.5 for speaking Spanish. In fact, only the returns to speaking German remain statistically significant in this regression. The results indicate that those who speak languages known by a smaller number of people obtain higher rewards in the labor market.[14]

We then check whether the returns to speaking foreign languages are homogeneous across the population. In particular, we analyze whether the returns differ for groups with different characteristics. We examine differences by gender, race, grades in college and graduate degree, by the percentage of Hispanics in the state of residence (for those whose second language is Spanish), and by occupation. The results (not reported in this version of the paper) reveal that speaking Spanish pays off less in states in which larger shares of the population are of Hispanic origin (probably because of a supply effect). By occupation, individuals in business support, management positions, or occupations related with science and technology are the ones who are more highly rewarded in the labor market for their foreign language skills. We do not find any statistically significant differences for the other individual characteristics analyzed.

These results should be interpreted with caution. In our regressions, we include a complete set of controls for ability (GPA and SAT scores), marital status, college quality, and parental background, to try to reduce the concern of omitted variable biases. Still, we acknowledge the possibility that unobservables correlated with wages and the ability to speak foreign languages may bias our OLS estimates. We therefore use information on changes in the ability to speak a second language to examine whether the estimates from our baseline earnings equations simply capture an ability bias. Our strategy is the following. If learning a second language requires cognitive skills that are not captured in our baseline specifications, the coefficient on the second language indicator may be capturing the effect on earnings of some of these. We can exploit the differences between the responses regarding second language knowledge between the 1993 and 1997 interviews to examine whether people with the necessary ability to learn a second language, but who could not speak it in 1997, earn the same wage premium as second language speakers.[15]

[14] This is consistent with anecdotal evidence. See, for example, *Wall Street Journal*, November 13, 1999: "Where bilingual workers are in short supply, the employees may command a premium or land a job they wouldn't otherwise have gotten."

[15] The transition matrix in and out of speaking a second language between 1993 and 1997 is the following:

	Speaks FL in 1993	Does not speak FL in 1993	Obs:
Speaks FL in 1997	2621	677	3298
Does not speak FL in 1997	1103	4947	6050
Observations:	3724	5624	9348

In Table 4 we replicate the previous OLS regressions, but now we add a dummy that equals one if the individual can speak a second language in 1993 but not in 1997. If the cognitive ability necessary to learn a second language is the only determinant of earnings, we should see a similar coefficient for those who could speak a second language in 1997 and those who were able to speak it in 1993. In fact, the coefficient on the indicator for speaking a second language in 1993 only is not very different from zero and is statistically insignificant. We proceed with the same estimation for the subsamples of native English speakers and of native speakers who are American citizens and whose parents were born in the U.S. We do so in order to separate the dynamics of those who learn a second language from those of language minorities because the different composition of the two groups could, in principle, make the interpretation of the results difficult. The results confirm that the returns to a second language are not completely driven by an ability bias.

These estimates exploit the fact that some individuals decide to learn a language and do not subsequently exert an effort in keeping it alive. The returns of speaking the language for the latter may have been very low. An unbiased estimate of the ex-ante returns to speaking a second language, taking into account that some forget it, is given by the difference between the coefficient of those who speak minus the coefficient of those who forgot, multiplied by the fraction of people who speak now out of those who ever spoke a second language.[16] Such estimates yield an approximate return of 2 percent.

5.2 Propensity Score Methods

If the returns to speaking a second language vary depending on individual characteristics, or if the people who speak and do not speak a second language are very different, as we saw in Table 1, our estimates of the returns to speaking a second language may be biased because of the linearity imposed in OLS regressions. To explore how sensitive our results are to the linearity assumption, we opt for propensity score methods to give flexibility to the functional form of the effect on earnings of speaking a second language across groups.

The propensity score is the probability of being assigned to a treatment, conditional on a set of covariates. The treatment, in our setting, is speaking a foreign language. The estimated propensity score allows us to control for differences between treatment and control groups when the treatment is not randomly assigned, the number of predetermined variables is high, and the groups are not very similar. It conveniently summarizes all individual characteristics in a single variable in the

[16] See Appendix B for an extensive discussion.

unit interval. Under selection on observables, conditioning on the propensity score is enough to have independence between the treatment indicator and the potential outcomes (Rosenbaum and Rubin, 1983).

While the assumption of selection on observables may seem strong in this setting, Dehejia and Wahba (1999) show that propensity score methods yield estimates of treatment effects that are closer to experimental benchmark estimates than traditional econometric techniques for non-experimental data.[17] Thus, although the propensity score is unknown and has to be estimated, inference for average treatment effects seems to be less sensitive to specifications of the propensity score than to the specification of the conditional expectation of potential outcomes implicit in OLS regressions.

We first estimate the propensity score for our sample using a logistic probability model. Table A.3 in the Appendix presents the coefficients from this estimation.[18] Conditional on the propensity score, the covariates should be independent of the assignment to the treatment, i.e., of whether the individuals speak a foreign language. We grouped the observations into strata defined on the propensity score and checked whether the covariates were balanced within each stratum. We did not find significant differences in the distribution of the covariates within these groups. After estimating the propensity score, we use matching, weighting on the propensity score, and stratification to produce alternative estimates of the returns to speaking a second language.

With stratification, we begin by sorting the observations according to their estimated propensity score. Panel A in Table 5 shows the differences in average earnings between treatment and control units within the propensity score deciles. We estimate the treatment effect by adding the within-strata differences in earnings and weighting by the number of treated observations in each decile. The estimated effect of speaking a second language is 1.2 percent for those who speak (the selected average treatment effect, or *SATE*). We also obtain an alternative estimate after adjusting for covariates, to eliminate any possible remaining within-block differences. One would expect this

[17] These authors compare a randomized evaluation of the National Supported Work (NSW) Demonstration with non-experimental comparison units from survey data sets, following Lalonde's analysis of this issue in 1986. They conclude that "there may be important unobservable covariates for which the propensity score method cannot account. However, [...], there are substantial rewards in exploring first the information contained in the variables that are observed. Propensity score methods can offer both a diagnostic on the quality of the comparison and a means to estimate the treatment impact."

[18] Note that we are conditioning on pre-treatment variables only (gender, race and family background). Rosenbaum (1984) shows that if the treatment is only ignorable given some *unobserved* pretreatment variables, propensity score methods may yield biased estimates if they condition on observed pretreatment variables. If this is the case, adjustment for pretreatment and some posttreatment characteristics may yield unbiased estimates. We therefore explored alternative specifications of the logistic model that included post-treatment individual characteristics. In results not reported here, we found that the estimates of the effect of the treatment on earnings remained unchanged.

adjustment to have little effect if the covariates are well balanced within the strata. Indeed, our estimate of the effect of speaking a second language on wages increases only slightly, up to 2 percent.

In the second method based on the propensity score, we match each individual who speaks a foreign language with the comparison individual with the closest propensity score.[19] We then drop from our sample the observations that correspond to individuals who do not speak a second language and were not matched to "treated" individuals. Table A.3 in the Appendix presents the characteristics of the matched sample for those individuals who speak a foreign language and those who do not. For this matched sample, we estimate a return *(SATE)* to speaking a second language of 2.8 percent (Panel B, Table 5). We obtain an adjusted estimate of the returns to speaking a second language including covariates in the regression, to account for any remaining observable differences between treated and control individuals. The point estimate increases slightly, up to 4 percent.

A final method of exploiting the propensity score is through weighting by the inverse of the probability of receiving the treatment (Horvitz and Thompson, 1952). Using this technique, the point estimate remains close to those obtained with the previous methods, 1.1 percent. Standard errors increase relative to other methods (.098) and make it impossible to reject a zero return to speaking a foreign language.[20] This finding is not surprising, though, as this is a relatively inefficient estimating method.[21]

All propensity score methods confirm our findings from the OLS regressions: speaking a foreign language is a valuable skill in the job market. Its returns are around 2-4 percent. While the identification conditions that OLS and propensity score methods require for the estimates to be unbiased are the same, the approach that we explore in this section helps us construct estimates of the conditional expectation of potential outcomes without the functional form restrictions that OLS regression imposes. The fact that all the estimates point in the same direction and are consistent across specifications and methods is reassuring and confirms our previous conclusion that speaking a second language is associated with a wage premium in the labor market.

[19] We perform matching with replacement, that is, allowing the same observation for a control individual to provide the match for several treatment units. See Dehejia and Wahba (1999).

[20] As the distribution of standard errors is not analytically tractable, we obtain standard errors using bootstrapping. Parameters were estimated 200 times to obtain an estimate of their variance. Each of the samples used for the estimation is constructed sampling the original sample with repetition.

[21] See Hirano, Imbens, and Ridder (2000).

5.3 Panel Data Methods

In this section we use the information on earnings and self-reported foreign language skills from two consecutive surveys to examine if the observed wage profiles of those who learn and forget a second language between the two surveys are consistent with the results obtained with OLS and propensity score methods. B&B contains information on second language ability in 1993 and 1997 and on wages in its 1994 and 1997 waves. The 1994 interview takes place when college graduates have just started their first job after graduation.

We exploit the longitudinal aspect of the data set to address the omitted ability problem inherent in the previous estimation methods. Consider the following explanatory model for the logarithm of wages:

$$\ln w_{it} = \alpha_i + \gamma SL_{it} + \beta' X_{it} + \rho t + \epsilon_{it}$$

where i and t are subscripts for individuals and time, respectively. SL_{it} is a dummy that equals 1 if individual i speaks a second language at time t and X_{it} is a vector of explanatory variables. γ is the coefficient of interest. ρ can be interpreted as a time effect, common to all individuals. α_i is an individual fixed effect. One can broadly interpret this coefficient as individual unobserved constant heterogeneity like ability. The problem with our OLS specification is that we cannot observe α_i. If SL_{it} and α_i are correlated, the coefficient on the second language indicator may be partially capturing the effect of ability on earnings. This suggests using the differences between the values of the variables in the two periods for which we have information:

$$(\ln w_{i1} - \ln w_{i0}) = \gamma(SL_{i1} - SL_{i0}) + \beta'(X_{i1} - X_{i0}) + \rho + (\epsilon_{i1} - \epsilon_{i0})$$

Table 6 shows the results of the estimation of this model. Note that only those variables that experience any change between both surveys appear in differenced form. The results that we obtain, in column one, are very similar to our OLS estimates. The wages of people who learn (forget) a second language tend to increase by almost 2 percent more (less) than other individuals.

One may worry that people who have the ability to learn a second language differ in their wage level and also in their wage profile over time. If this were the case, our estimate of γ in this first specification would be capturing the steeper wage profile over time for higher ability individuals. One would then expect that those who speak a second language in both periods (call them *always speakers*, AS) should also experience higher wage increases. In column 2 of Table 6 we include an AS indicator to account for this possibility. As the results show, people who spoke a second language in both 1993 and 1997 do not experience higher wage increases over time. In column 3 we repeat this exercise for the subsample of individuals who spoke English at home while they were growing up, are American citizens, and have American-born parents. Again, we want to render the groups of those who learn a second language later in life and the AS as similar as possible. Our estimate of γ varies little across these specifications. These results thus suggest that the intrinsic knowledge of a second language is the explanation for the wage hike captured in the panel estimate of γ. As in our OLS regressions, the results are consistent with a labor market return of around 2 percent of learning a second language.

Another possibility is that the wage profile (the steepness of wages with respect to seniority) may differ according to the observable characteristics of the individuals:

$$\ln w_{it} = \alpha_i + \gamma SL_{it} + \beta' X_{it} + \rho t + \delta_0' X_{i0}(1-t) + \delta_1' X_{i0} t + \epsilon_{it}$$

If the characteristics of the individuals who learn and forget a second language between the two surveys are different, our estimate of γ would capture only such differences. Taking differences:

$$\Delta \ln w_{it} = \gamma \Delta SL_{it} + \beta' \Delta X_{it} + \rho + \delta' X_{i0} + \epsilon_{it}$$

where δ is $\delta_1 - \delta_0$. In columns 4 and 5 of Table 6 we estimate this model. We control for the 1994 values of the individual characteristics that we used in our baseline specification as regressors. Column 4 includes all individuals in the sample and column 5 is limited to those English-speaking natives, third-generation Americans. The results are in line with those in previous specifications.

There are two caveats in the interpretation of the previous estimates. The first is the common problem of false transitions. Many people who spoke a second language in 1993 and declare they couldn't speak it in 1997 may actually be capable of doing so after some short training. Some of them may not have been capable of speaking it in the first period. People who declare they have learned a second language may have actually spoken it in the first wave but were not sure about their answer. It is likely that the measurement error in the question "Do you have conversational knowledge of languages other than English?" is larger for the group who do not give a consistent answer in 1993 and 1997 than for the group that did not change its response. This amounts to bigger measurement error in the variable capturing whether the individual's knowledge of a foreign language changes between 1993 and 1997. This should bias our estimates of γ downward.

A second problem arises from the selection of individuals into learning and forgetting a second language. If our results are not driven by ability bias, learning a second language is, indeed, a profitable investment. Thus, we should expect people with the lower returns to speaking a second language not to make the effort toward its maintenance. Conversely, those learning a second language after college are most likely the people with higher labor market returns to speaking it. These facts make the interpretation of our results in Table 6 difficult. More specifically, the returns for those who learn a second language should be higher than the returns for those who forget it. Our specification in Table 6 imposes symmetry in the returns for both groups. In Table 7 we replicate the structure of the previous regressions, but we now allow for different returns for the two groups. Note that the returns for those who learn and those who forget a second language can be interpreted as a higher and lower bound for the returns to learning a second language. As we expected, people who learn a second language between the two B&B surveys have a higher estimated return (almost 4 percent) than people who report having forgotten it (0 percent). The results confirm that previous estimates capture an actual labor market value of speaking a second language, and give us an interval between 0 percent to 4 percent for its return.

5.4 Instrumental Variable Methods

We finally exploit exogenous variation in the decision to learn a foreign language. A consistent estimate of the returns to speaking a foreign language can be obtained if there is a variable that affects whether an individual speaks a second language but which does not directly affect earnings. Note that the interpretation of the IV estimates, however, will differ slightly from the coefficients we have previously estimated. If the instruments are valid, and if the returns to speaking a foreign language vary across the population, we will obtain a consistent estimate of the Local Average Treatment Effect *(LATE)*.[22] *LATE* is the average return for the population of *compliers*, that is, those individuals who learn a second language because they are required to do so, and who would not have learned it if not required. In principle, the economic return might be different for this group of people than for the average individual in the population (the average treatment effect, ATE) or the average individual who speaks a foreign language (SATE).

We use high school and college graduation requirements as determinants of whether an individual speaks a foreign language that can be excluded from the earnings equation. In this section we first discuss the validity of the instruments and then present our IV estimates of the returns to speaking a second language.

[22] See Angrist and Imbens (1994).

5.4.1 Are Graduation Requirements Legitimate Instruments?

Most of the individuals in our sample graduated from high school in 1989. At the time, several states required high school graduates to complete a certain number of courses in foreign languages (District of Columbia, Rhode Island, and Texas) or included foreign languages among several elective subjects students could choose among in order to complete the necessary number of credits to obtain their high school diploma (California, New Hampshire, Oklahoma, Oregon, Virginia and West Virginia). Similarly, some colleges require all or some of their students to satisfy foreign language requirements, either by taking a certain number of courses or by demonstrating proficiency in a foreign language test. Table A.5 in the Appendix lists the states and their graduation requirements.

Unfortunately, B&B does not report information on the states where the individuals in the sample completed high school. We have information, however, on the state where their parents lived in 1993. We therefore assume that the parents have not moved to other states since their children graduated from high school, so that the individuals we observe completed high school in the states where their parents lived in 1993.[23] In Panel A of Appendix Table A.6 we compare individuals whose parents live in states that had foreign language requirements to graduate from high school graduation with those that do not . We observe some differences between these two groups' hourly earnings, the likelihood of speaking foreign languages, of attending a public college (higher in states with requirements), and of being married (lower in states with requirements). States with foreign language requirements had slightly higher income per capita in 1989, when the average individual in our sample was graduating from high school.

We perform the same comparison between individuals who were subject to foreign language requirements in college and individuals who were not (Table A.6, Panel B.1). Some differences arise between these two groups in age, experience, marital status, and race. Moreover, colleges with requirements are located in states with lower per capita income. The proportion of colleges with requirements varies with quality, and the lowest proportion of colleges with requirements is found among the least competitive ones. The relation between quality and requirements is non-monotonic, however: there are more colleges with requirements in the "very competitive" category than among the "most competitive" ones, for example. Public colleges emphasize foreign language requirements more than private ones.

[23] It is important to note that high school graduation requirements in 1989 are likely to be relevant for the individuals who actually completed high school in that year. As the summary statistics for our sample revealed, there is some dispersion in the age distribution. The results in this section remain unchanged, though, if we focus on a subsample of individuals born between 1968 and 1972.

These graduation requirements will be legitimate instruments if they affect an individual's decision to learn a foreign language but have no direct effect on earnings. The observed differences between individuals affected or not by the requirements suggest that we should include these variables as controls in our regressions. To control for the possibility that families that place a strong emphasis on education choose to live in states with stronger high school graduation requirements, we include the set of requirements in other subjects as controls in our regressions.[24] We think it is safe to assume that foreign language requirements *per se* do not drive the residential decisions of families in search of better school systems. The same reasoning drives us to include a complete set of college quality indicators, to reduce the concern that higher quality colleges impose stronger foreign language requirements on their students. Once we include these controls, we believe that graduation requirements are legitimate instruments and that they are not correlated with unobservable determinants of individuals' earnings. We find it hard to argue that these requirements drive the individual's or her family's decision to attend a specific college or to live in a certain state.

5.4.2 Instrumental Variables Estimates

We use high school and college graduation requirements and their interactions as our instruments. We show the results from the first stage estimation in Table 8. Both types of graduation requirements significantly increase the probability of speaking a foreign language. High school graduation requirements are the most effective, increasing the probability that the individuals speak a foreign language by more than 18 percent.

Table 9 presents the estimates of the labor market returns to speaking a foreign language using college and high school graduation requirements as instruments for the ability to speak a foreign language. We estimate alternative specifications, including in some of them controls for other high school requirements and college quality. Regardless of the specification, the conclusions are similar. The use of graduation requirements as exogenous determinants of whether an individual speaks a foreign language yields estimates of the returns to speaking a foreign language between 14 and 30 percent, much higher than the OLS estimates. Standard errors are large, as well, so that we can not reject the hypothesis that the difference between the OLS and the IV estimators is zero at conventional significance levels.[25]

[24] The number of semesters in English, Social Studies, Math, and Science. See Table A.4 in the Appendix for a comprehensive list of state requirements.

[25] We experimented with alternative sets of instrumental variables in unreported regressions. One particular concern is that college graduation requirements may reflect unobserved college quality and selectivity. It is thus reassuring that when we focus only on state high school graduation requirements our results do not change. We also

We estimate alternative IV models to test the robustness of our findings. As we did for the OLS estimates, we restrict the sample to individuals who spoke English at home while growing up (column 2), to American citizens who spoke English at home and whose parents were born in the U.S. (column 3), and to full-time workers who spoke English at home (column 4). Again, the point estimates remain stable across specifications, although they are large and very imprecise.

How can we interpret the differences between the returns to speaking a second language that OLS regressions and propensity score methods suggest and our IV estimates? If we believe that the "true" rate of return is homogeneous across the population, the IV estimates that rely on exogenous sources of variation in the ability to speak a second language indicate that conventional OLS estimates are downward biased. However, our preferred interpretation is the following: if the rate of return to speaking a foreign language does vary across the population, and the population that is affected by the graduation requirements (the *compliers* that we previously mentioned) have high returns to learning a foreign language, the IV estimate, which depends on the marginal return for this subset of the population, will be relatively high.[26]

6 Conclusions

The popular press has repeatedly presented anecdotal evidence suggesting that speaking a foreign language is a valuable skill. However, little systematic research has addressed this issue. In this paper we have explored this question using a variety of empirical strategies. While none of them provides an ideal approach to the problem, all of them point in the same direction: speaking a foreign language is rewarded in the labor market. The earnings of those who speak a foreign language are higher than the earnings of those who don't. This earnings differential does not seem to be driven by unobservable differences between these two groups of individuals, as our IV estimates show. The returns are not homogeneous across the population but are nevertheless positive and significant. Learning a foreign language is an investment with positive monetary returns for the average college graduate in the U.S.

This result could be valuable for policy-making purposes. In particular, frequent debates frame the decisions to define high school curricula. Our results confirm that speaking a second language is

included the graduation requirements by state colleges, the changes in high school graduation requirements over time, etc. The estimates remain around 20 percent and very imprecise.

[26] A possible explanation for the high point estimates that we obtain in the IV regressions is that the second language indicator is measured with nonclassical measurement error, which would bias the magnitude of IV coefficients upwards (Kane *et al.,* 1999). Another possibility is that we may also be introducing some contaminating variation if the instruments reflect unobserved individual and/or family background characteristics.

rewarded in the labor market with higher average earnings. It maybe therefore advisable to require high school students to show proficiency in a foreign language before they graduate.

We believe that further research should assess the reasons for the positive returns to speaking a foreign language. It would be interesting to examine if speaking a foreign language is used as a signal for other abilities, for example. Is this skill valued per se, or only when it is actually necessary on the job? For example, is it more valuable in those occupations and industries in which foreign relations or international trade play an important role? The extension of the analysis to other population groups in the U.S. or to individuals in other countries could show whether the returns that we estimate in this paper extend to other populations. All these questions deserve future exploration.

References

[1] Altonji, Joseph G. (1995) "The Effects of High School Curriculum on Education and Labor Market Outcomes," *Journal of Human Resources*, 30(3): 409-38.

[2] Angrist, J., and Guido Imbens (1994) "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2): 467-476.

[3] Angrist, J., and Alan B. Krueger (1999) "Empirical Strategies in Labor Economics," in *Handbook of Labor Economics*, Volume 3A. O.Ashenfelter, and David Card, eds. Amsterdam; New York and Oxford: Elsevier Science, North-Holland.

[4] Becker, G. (1964) *Human Capital*. New York: Columbia University Press.

[5] Brod, R., and M. Lapointe (1989) "The Modern Language Association Survey of Foreign Language Entrance and Degree Requirements, 1987-88," *ADFL Bulletin*, 20(2).

[6] Carliner, G. (1995) "The Language Ability of US immigrants: Assimilation and Cohort Effects," NBER Working Paper #5222.

[7] Chiswick, B.R. (1997) "Hebrew Language Usage: Determinants and Effects on Earnings Among Immigrants in Israel," *Journal of Population Economics*, 11(2): 253-271.

[8] Chiswick, B.R., and P.W. Miller (1997a) "Language Skill Definition: A Study of Legalized Aliens," Mimeo.

[9] Chiswick, B.R., and P.W. Miller (1997b) "Language and Earnings Among Immigrants in the United States," Mimeo.

[10] Chiswick, B.R., and P.W. Miller (1998) "English Language Fluency Among Immigrants in the United States," *Research in Labor Economics*, 17: 151-200.

[11] Chiswick, B.R., and P.W. Miller (2000) "The Complementarity of Language and Other Human Capital: Immigrant Earnings in Canada," Center for the Study of the Economy and the State,Graduate School of Business, University of Chicago, Working Paper No. 148.

[12] Christofides, L.N. and Swidinsky, R. (1998) "Bilingualism and Earnings: A Study Based on 1971, 1981 and 1991 Census Data," in *New Canadian Perspectives. Economic Approaches to Language and Bilingualism*, A.Breton, ed.

[13] Cook, V. (1997) "The Consequences of Bilingualism for Cognitive Processing," in *Tutorials in Bilingualism: Psycholinguistic Perspectives*, A.M. De Groot, and J.F. Kroll, eds. Mahwah, New Jersey: Lawrence Erlbaum Associates.

[14] Cooper, T.C. (1987) "Foreign Language Study and SAT-Verbal Scores," *The Modern Language Journal*, 71 (4).

[15] Dehejia, R.V., and Sadek Wahba (1999) "Causal Effects in Non-experimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94(448): 1053-1062.

[16] Grenier, G., and F. Vaillancourt (1983) "An Economic Perspective on Learning a Second Language," *Journal of Multilingual and Multicultural Development*, 4(6).

[17] Griliches, Z. (1977) "Estimating the Returns to Schooling: Some Econometric Problems," *Econometrica*, 45(1): 1-22.

[18] Heckman, James (1980) "Sample Selection Bias as a Specification Error," in *Female Labor Supply*, J. Smith, ed. Princeton NJ: Princeton University Press.

[19] Hirano, K., Imbens, G.W. and Geert Ridder (2000) "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," NBER Working Paper No. T0251.

[20] Hoffman, C. (1991) *An Introduction to Bilingualism.* London: Longman.

[21] Horvitz, D.G. and D.J.Thompson (1952) "A Generalization of Sampling Without Replacement from a Finite Universe." *Journal of the American Stastistical Association*, 47: 663-685

[22] Hoxby, C. M., and B. T. Long (1999) "Explaining Rising Income and Wage Inequality Among the College Educated," NBER Working Paper No. 6873.

[23] Kane, Thomas J., Cecilia E. Rouse, and Douglas Staiger (1999) "Estimating Returns to Schooling When Schooling Is Misreported," NBER Working Paper.

[24] Lalonde, R. (1986) "Evaluating the Econometric Evaluations of Training Programs," *American Economic Review* 76(4): 604-620.

[25] Lopez, M.H. (1999) "Does Speaking a Second Language Affect Labor Market Outcomes? Evidence from National Adult Literacy Survey of 1992," mimeo: School of Public Affairs, University of Maryland.

[26] Mincer, J. (1974) *Schooling, Experience and Earnings*. New York: Columbia University Press.

[27] Olsen, S., and L.K. Brown (1992) "The Relationship Between High School Study of Foreign Language and ACT English and Mathematics Performance," *ADFL Bulletin*, 23(3):47-50.

[28] Pendakur, K., and R.Pendakur (1998) "Speak and Ye Shall Receive: Language Knowledge as Human Capital" in *New Canadian Perspectives. Official Languages in Canada: Changing the Language Landscape*, A.Breton, ed. Ottawa : Canadian Heritage.

[29] Ridler, N.B. and S.Pons-Ridler (1984) "Language Economics: ACase Study of French," *Journal of Multilingual and Multicultural Development*, 5,1: 57-63.

[30] Rosenbaum, P. (1984) "The Consequences of Adjustment for a Concomitant Variable That Has Been Affected by the Treatment," *Journal of the Royal Statistical Society*, 147(5).

[31] Rosenbaum, P., and D. Rubin (1983) "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70: 41-55.

[32] Rosenbaum, P., and D. Rubin (1984) "Reducing Bias in Observational Studies Using the Subclassification on the Propensity Score," *Journal of the American Statistical Association*, 79: 516-524.

[33] Rubin, D. (1974) "Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies," *Journal of Educational Psychology*, 66: 688-701.

[34] Rubin, D. (1977) "Assignment to a Treatment Group on the Basis of a Covariate," *Journal of Educational Statistics*, 2:1-26.

[35] Rubin, D. (1978) “Bayesian Inference for Causal Effects: The Role of Randomization,” *The Annals of Statistics*, 6: 34-58.

[36] Shapiro, D.M., and M. Stelcner (1997) “Language and Earnings in Quebec: Trends over Twenty Years, 1970-1990,” *Canadian Public Policy*, 23(2): 115-140.

A Data Appendix

A.1 Quality of Schooling Variables

This is a set of dummies that indicate the quality of the college. The classification was provided by Bridget Terry Long and is defined in detail in Hoxby and Long (1999). Colleges were grouped according to criteria taken from the 1997 edition of Barron's *Profiles of American Colleges*. These criteria include raw test scores, percentage of the class scoring over a certain level, and high school rank. Schools that had not been ranked were categorized according to these criteria. These have remained fairly constant over the last 15 years, and thus the use of the 1997 list is valid. Scores were reported as nonrecentered. The SAT scores were converted into percentiles.

Quality Groupings Criteria

Category	G.P.A.	S.A.T.	A.C.T.	Sample Schools
Most Competitive	3.75 - 4.50	1280 - 1600	31 - 36	Princeton, Harvard, UVA
Highly Competitive	3.40 - 3.74	1120 - 1279	26 - 30	UC Berkeley, University of Michigan
Very Competitive	3.00 - 3.39	1000 - 1119	21 - 25	University of Connecticut
Competitive	2.50 - 2.99	850 - 999	16 - 20	University of Massachusetts
Less Competitive	2.00 - 2.49	650 - 849	12 - 15	University of Mississippi
Non Competitive	1.00 - 1.99	649 - 400	4 - 11	Ohio State University - Mansfield Campus

Source: Barron's Educational Series, *Profiles of American Colleges*, 21st Edition (1997).

Notes: The categories were created referring to Barron's criteria for school groupings, which are based on the characteristics of the freshman class entering 1995-96 (before SAT scores were recentered).

B A Simple Model

The following simple model illustrates the relation between ability and the returns to learning a second language. It provides a justification for the strategy followed in the regressions in Table 7. Assume that the decision on learning a second language (SL hereafter), which involves a period of training, is undertaken before full participation in the labor market. Assume, furthermore, that there are two kinds of individuals. A fraction α of the population has low costs of learning an SL, C_L (which corresponds to high ability). The rest of the population has costs C_H. Ability is not observable by the econometrician. Assume that if the second language is learned, it will turn out to be useful (producing extra earnings V_U) with probability p and turn out not to be useful in the labor market with probability $(1-p)$. Assume that there are no consumption advantages from speaking a foreign language. Labor market earnings are:

$$w_i = \begin{cases} V_u + A - \beta C_i, & \text{if the individual speaks an SL and the SL realization is positive} \\ A - \beta C_i, & \text{in all other cases} \end{cases}$$

High ability individuals will decide to learn the second language if:

$$p \cdot V_U \geq C_L$$

For low ability individuals, the condition is:

$$p \cdot V_U \geq C_H$$

Consider the case when:

$$C_H \geq p \cdot V_U \geq C_L$$

Then, all high ability individuals learn a second language and low ability individuals do not.

Assume that individuals can maintain their foreign language human capital after the realization of the labor market shock with a very small cost ε. High ability individuals for which the foreign language turned out to be useful will decide to maintain their stock of foreign language capital. High ability individuals for which the second language did not turn to be useful will decide not to maintain it. Ex-post, only individuals with high ability and a positive realization will speak a second language. The pool of those who do not speak an SL will be formed by low ability individuals and high ability individuals with negative labor market realizations.

We know that the expected return of learning a second language, corresponding to a complete randomization of the second language treatment, is equal to $p \cdot V_U$. We may try to estimate the returns to learning a second language using the difference between the wages of those who speak an SL and the rest $(w_s - w_{ns})$. Let T be an indicator that takes value one if the individual speaks a second language and zero otherwise. The expectation of this estimator is:

$$\begin{aligned} E\left(w_s - w_{ns}\right) &= E(w|SL=1) - E(w|SL=0) = \\ &= [V_u + A - \beta C_L] - \left[\frac{(1-p)\alpha}{1-\alpha p} \cdot (A - \beta C_L) + \frac{(1-\alpha)}{1-\alpha p} \cdot (A - \beta C_H)\right] = \\ &= V_u + \frac{\beta(1-\alpha)}{(1-\alpha p)} \cdot [C_H - C_L] > V_u > pV_u \end{aligned}$$

This estimate is too high for two reasons. First, there is an ability bias: individuals who speak a second language have higher average unmeasured ability. Second, there is a problem of selection by earnings: those who speak a second language ex-post tend to have greater returns than the rest.

These biases suggest the use of the alternative estimator: $p \cdot (w_s - w_{forgot})$, where w_{forgot} is the wage for those individuals who learned the SL and did not invest in its maintenance ex-post. This is an unbiased estimator of the ex-ante return to speaking an SL:

$$E\left(w_s - w_{forgot}\right) = E(w|SL = 1) - E(w|forgot) = [V_u + A - \beta C_L] - [A - \beta C_L] = V_u$$

Multiplying by the fraction of people who speak a second language out of the total who ever spoke it yields the ex-ante expected returns to learning a second language. From column (1) in Table 4 we can derive:

$$(w_s - w_{forgot}) = .029 - .002 = .027$$

The fraction of people who speak a second language out of the fraction who ever spoke it is equal to the number of people who speak divided by the number of people who speak plus the number of people who forgot:

$$p = \frac{N_S}{N_S + N_{forgot}} = \frac{3298}{3298 + 1103} = 0.75$$

Thus, our estimate of the ex-ante expected returns of learning an SL is:

$$p \cdot (w_s - w_{forgot}) = 0.020$$

Note how this is a conservative estimate of the expected returns to learning an SL, as we are assuming that the returns for those who forgot the second language is zero. Furthermore, note that the fact that the coefficient for those who forgot a second language is close to zero suggests that the ability bias is small.

Table 1:
Comparison of Individuals Who Speak and Do Not Speak Foreign Languages, 1997 Survey

Panel A: Descriptive Statistics

Variable:	Speak	Not Speak	Difference
Working?	.955	.967	-.012***
	(.207)	(.177)	(.004)
Log Hourly Wage	2.562	2.544	.018*
	(.432)	(.414)	(.009)
Speaks F.L.	1	0	–
	–	–	
Age	29.611	30.356	-.744***
	(6.083)	(6.755)	(.153)
Experience (Months)	42.185	43.842	-1.657***
	(12.342)	(11.011)	(.270)
Married	.244	.337	-.093***
	(.429)	(.472)	(.010)
Black	.090	.014	.075***
	(.286)	(.119)	(.004)
Male	.445	.44	-.003
	(.498)	(.497)	(.011)
Normalized GPA	3.062	3.063	-.001
	(.492)	(.502)	(.011)
Public College	.653	.689	-.035***
	(.475)	(.462)	(.011)
MBA	.023	.031	-.008**
	(.149)	(.175)	(.003)
Ph.D.	.043	.021	.021***
	(.204)	(.146)	(.004)
Other Master's	.269	.249	.020**
	(.443)	(.432)	(.010)
Observations:	2756	5184	7940
	(34.71%)	(65.28%)	

Panel B: SAT Scores

Quartile:	Speak	Not Speak	Difference
1	.170	.203	-.033***
	(.376)	(.403)	(.009)
2	.202	.227	-.025***
	(.401)	(.419)	(.009)
3	.217	.198	.018**
	(.412)	(.398)	(.009)
4	.219	.165	.054***
	(.413)	(.371)	(.009)
Obs:	2756	5184	7940

Panel C: Regional Distribution, State of Residence

Region:	% Speak	Observations
New England	.41	485
Middle Atlantic	.38	999
East North Central	.26	1231
West North Central	.27	689
South Atlantic	.35	1658
East South Central	.26	541
West South Central	.32	840
Mountain	.34	544
Pacific	.41	1028

Notes:

1. Subsample of individuals who have hourly wages above $2.8 (1 percentile) and below $42.3 (99 percentile), who answer the question on whether they speak a foreign language and with complete data on age, experience, gender, marital status, race, state of residence, college GPA, and type of college attended.
2. Sample means weighted using sample weights.
3. *** statistically significant at the 1% level, ** at the 5% level, * at the 10% level.
4. New England: CT, ME, MA, NH, RI, VT; Middle Atlantic: NJ, NY, PA; East North Central: IN, IL, MI, OH, WI; West North Central: IA, KS, MN, MO, NB, ND, SD; South Atlantic: DE, DC, FL, GA, MD, NC, SC, VA, WV; East South Central: AL, KY, MS, TN; West South Central: AR, LA, OK, TX; Mountain: AZ, CO, ID, NM, MT, UT, NV, WY; Pacific: AK, CA, HI, OR, WA.

Table 2: Descriptive Statistics, Individuals Who Speak a Foreign Language, 1997 Survey

Panel A: Do They Speak More Than One Language?

Sample:	All	Women	Men	Difference
Speaks 1 F.L.	.840	.841	.838	-.0006
	(.366)	(.365)	(.367)	(.014)
Speaks 2 F.L.	.135	.135	.134	-.001
	(.342)	(.342)	(.341)	(.013)
Speaks 3 F.L.	.018	.019	.018	-.001
	(.136)	(.137)	(.134)	(.005)
Speaks >3F.L.	.005	.003	.007	.004
	(.076)	(.056)	(.088)	(.010)
Observations:	2756	1530	1226	2756

Panel B: What Languages Do They Speak?

Sample:	All	Women	Men	Difference
Spanish	.58	.586	.589	.002
	(.49)	(.492)	(.492)	(.019)
French	.23	.272	.181	-.090***
	(.42)	(.445)	(.385)	(.016)
German	.11	.081	.149	.067***
	(.31)	(.273)	(.356)	(.012)
Italian	.03	.043	.023	-.020***
	(.18)	(.204)	(.150)	(.007)
Russian	.01	.012	.021	.008*
	(.12)	(.112)	(.145)	(.005)
E/SE Asian	.04	.026	.062	.035***
	(.20)	(.161)	(.242)	(.007)
Chinese	.02	.024	.027	.002
	(.15)	(.156)	(.163)	(.006)
Obs.:	2756	1530	1226	2756

Panel C: Regional Distribution

Region:	N. E.	Middle Atlantic	E.N Central	W. N. Central	South Atlantic	E. S. Central	W. S. Central	Mountain	Pacific
Spanish	.52	.53	.54	.45	.61	.67	.72	.65	.56
French	.34	.25	.22	.17	.28	.25	.15	.14	.20
German	.09	.10	.15	.16	.09	.08	.10	.08	.10
Italian	.09	.08	.01	.008	.03	.02	.006	.008	.01
Russian	.02	.01	.01	.008	.02	.006	.005	.04	.008
E/SE Asian	.008	.05	.04	.16	.01	.009	.01	.01	.06
Chinese	.03	.04	.01	.01	.01	.005	.02	.008	.04

Notes:

1. Subsample of individuals who have hourly wages above $2.8 (1 percentile) and below $42.3 (99 percentile), who answer YES to the question on whether they speak a foreign language and with complete data on age, experience, gender, marital status, race, state of residence, college GPA, and type of college attended.

2. See Notes 2, 3 and 4 to Table 1.

Table 3: OLS Estimates

Dependent Variable: Log Hourly Wage in 1997

	(1)	(2)	(3)	(4)
Sample:	All	Spoke English at Home	Spoke English at Home, American Citizen, Parents Born in US	Spoke English at Home, Works>35h.
Speaks F.L.	.028***	.022**	.022**	.019*
	(.01)	(.010)	(.010)	(.010)
Age	.030**	.035***	.032***	.031***
	(.006)	(.008)	(.006)	(.006)
Age^2	-.0003***	-.0003***	-.0003***	-.0003***
	(.00007)	(.0001)	(.00008)	(.00008)
Experience	.002	.004	.0016	-.004**
	(.002)	(.002)	(.002)	(.002)
Experience^2	-.00001	-.00004	-.00001	.00005*
	(.00003)	(.00003)	(.00003)	(.00003)
Male	.071***	.077***	.074***	.062***
	(.010)	(.010)	(.010)	(.010)
Married	.040***	.041***	.048***	.034***
	(.010)	(.011)	(.011)	(.011)
Black	.061***	.005	.020	-.001
	(.023)	(.032)	(.055)	(.031)
Log State Income	.515***	.449***	.494***	.442***
	(.034)	(.054)	(.037)	(.037)
Public College	.004	.006	-.002	-.007
	(.009)	(.011)	(.010)	(.010)
Normalized	.015*	.008	.011	-.001
College GPA	(.008)	(.009)	(.009)	(.009)
Parents' Education	Yes	Yes	Yes	Yes
College Quality	Yes	Yes	Yes	Yes
Graduate Degree	Yes	Yes	Yes	Yes
SAT-ACT Quartile	Yes	Yes	Yes	Yes
Major	Yes	Yes	Yes	Yes
Adj. R^2	.186	.182	.182	.198
Observations:	7940	7073	6629	5906

Notes:

1. Subsample of individuals who have hourly wages above $2.8 (1 percentile) and below $42.3 (99 percentile). Each regression is performed for the maximum number of observations for which all the covariates were non-missing.
2. Observations are weighted using sample weights.
3. Standard errors are reported in parenthesis.
4. Log State Income is the log of the per capita income in the state of residence in 1997 (BEA estimation).
5. Normalized GPA on a 0-4 scale for all respondents.
6. Parents' Education is a set of dummies that capture the education level of the individual's mother and father.
7. Majors is a detailed set of indicators for the student's major in college (100 categories).
8. Quality of College Dummies were provided by Bridget Terry Long. See the Appendix for a description, and Hoxby and Long, (1999) for details.
9. Graduate Degree dummies are three variables that indicate if the individual has a Ph.D., an MBA, or a Master's.
10. *** statistically significant at the 1% level, ** at the 5% level, * at the 10% level.

Table 4: A Test on Ability Bias

Dependent Variable: Log Hourly Wage in 1997

	(1)	(2)	(3)
Sample:	All	Spoke English at Home, American Citizen	Spoke English at Home, American Citizen, Parents Born in US.
Speaks FL in 1997	.029***	.024**	.024**
	(.010)	(.011)	(.011)
Spoke FL in 1993 only	.002	.007	.005
	(.014)	(.015)	(.015)
Age	.029***	.030***	.031***
	(.006)	(.008)	(.006)
Age^2	−.0003***	-.0003***	-.00031***
	(.0001)	(.00008)	(.00001)
Experience	.003	.003	.002
	(.002)	(.002)	(.002)
Experience^2	-.00002	-.00003	-.00002
	(.00003)	(.00003)	(.00003)
Male	.074***	.081***	.075***
	(.010)	(.010)	(.010)
Married	.039***	.050***	.049***
	(.011)	(.011)	(.011)
Black	.056**	.014	.019
	(.023)	(.032)	(.036)
Log State Income	.525***	.521***	.497***
	(.053)	(.036)	(.086)
Public College	.005	.004	-.0002
	(.010)	(.010)	(.010)
Normalized	.017*	.011	.010
College GPA	(.009)	(.009)	(.009)
College Quality	Yes	Yes	Yes
SAT-ACT Quartile	Yes	Yes	Yes
Major	Yes	Yes	Yes
Parents' Education	Yes	Yes	Yes
Graduate Degree	Yes	Yes	Yes
Adjusted R^2	.188	.182	.184
Observations:	7444	6632	6117

Notes:

1. Sample of individuals who have hourly wages above the 1st percentile and below the 99th percentile, who answer the question of whether they speak a foreign language in both surveys, and for whom the variables included in the regressions were complete.
2. Observations are weighted using sample weights.
3. Standard errors in parenthesis.
4. See notes to Table 3 for the variable definition.
5. *** statistically significant at the 1% level, ** at the 5% level, * at the 10% level.

Table 5: Estimated Impact of Speaking a Second Language
Propensity Score Methods

Panel A. Stratifying on the Score

	Not Adjusted			Adjusted		
	Log.Wage Diff.	#Treated	#Control	Log.Wage Diff.	#Treated	#Control
Decile 1	.005 (.037)	173	603	.009 (.035)	166	574
Decile 2	-.033 (.034)	207	585	.027 (.033)	195	556
Decile 3	-.029 (.034)	194	585	-.014 (.033)	181	559
Decile 4	.063* (.035)	194	575	.036 (.033)	188	550
Decile 5	-.006 (.032)	239	544	.036 (.031)	229	522
Decile 6	-.022 (.031)	271	505	.024 (.030)	259	480
Decile 7	-.025 (.032)	253	515	.000 (.031)	240	500
Decile 8	-.087*** (.031)	299	481	-.022 (.029)	287	459
Decile 9	.024 (.031)	314	434	.051* (.029)	304	414
Decile 10	.118*** (.034)	518	200	.036 (.036)	483	190
Observations:		2662	5027		2532	4804
	$\widehat{\alpha}_{SATE}^{Not\ Adj.}$	.012 (.011)		$\widehat{\alpha}_{SATE}^{Adj.}$	.020** (.010)	

B. Matching on the Score

	Unadjusted	Adjusted
Speaks FL	.028*** (.011)	.040*** (.011)
R^2	.001	.186
Observations: 5559		

Notes:

1. Sample of individuals who have hourly wages above the 1st percentile and below the 99th percentile, who answer the question of whether they speak a foreign language in both surveys, and for whom the variables included in the regressions were complete. In Panel B, the sample is further restricted to individuals who speak a foreign language and to those who don't with the closest propensity score.
2. Propensity scores are estimated using the logistic model presented in Table A.3.
3. Adjusted results include the set of controls presented in Table 3.
4. Standard errors in parenthesis.
5. *** statistically significant at the 1% level, ** at the 5% level, * at the 10% level.

Table 6: Exploiting Information from First Survey

Dependent Variable: Log Hourly Wage 1997-Log Hourly Wage 1994

Sample:	(1) All	(2) All	(3) Spoke English at Home, American Citizen, Parents Born in US	(4) All	(5) Spoke English at Home, American Citizen, Parents Born in US
ΔSpeaks FL	.017*	.017*	.022**	.012	.017*
	(.009)	(.009)	(.009)	(.009)	(.009)
Speaks F.L. both in 1997 and 1993	-	-.003	-0.011	.002	-.009
		(.009)	(.011)	(.010)	(.011)
ΔAge^2	.0012***	.0012***	.001***	.002**	.003***
	(.0001)	(.0001)	(.0001)	(.0008)	(.0009)
Experience	-.002	-.002	-.001	.0001	-.0003
	(.002)	(.001)	(.002)	(.0017)	(.0020)
Experience^2	.00005**	.00005**	.00005*	.00002	.00003
	(.00002)	(.00002)	(.00003)	(.00002)	(.00003)
ΔLog State Income	-.017	-.017	-.011	.015	.006
	(.051)	(.051)	(.055)	(.054)	(.059)
ΔMarried	.007	.007	.021**	.012	.026**
	(.009)	(.009)	(.010)	(.010)	(.011)
(Age in 1994)^2	-	-	-	-.00008	-.0002
				(.00007)	(.00008)
Male	-	-	-	.053***	.057***
				(.008)	(.009)
Married in 1994	-	-	-	.017	.013
				(.010)	(.011)
Black	-	-	-	-.037*	.001
				(.021)	(.054)
Log State Income 1994	-	-	-	.128***	.126**
				(.034)	(.037)
Public College	-	-	-	-.039***	-.036***
				(.009)	(.010)
Normalized College GPA	-	-	-	-.001	-.001
				(.003)	(.003)
College Quality	No	No	No	Yes	Yes
SAT-ACT Quartile	No	No	No	Yes	Yes
Major	No	No	No	Yes	Yes
Parents' Education	No	No	No	Yes	Yes
Graduate Degree	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	.032	.031	.028	.092	.092
Observations:	7686	7686	6329	7248	5977

Notes: See Notes to Table 3.

Table 7: Exploiting Information from First Survey
Allows Asymmetry between Returns to Learning and Forgetting
Dependent Variable: Log Hourly Wage 1997−Log Hourly Wage 1994

	(1)	(2)	(3)	(4)	(5)
Sample:	All	All	Spoke English at Home, American Citizen, Parents Born in US	All	Spoke English at Home, American Citizen, Parents Born in US
Speaks FL 1997 only	.037**	.038**	.028*	.035**	.026
	(.015)	(.015)	(.016)	(.016)	(.096)
Spoke FL 1993 only	-.003	-.003	-.017	.004	-.007
	(.012)	(.012)	(.013)	(.013)	(.013)
Speaks FL both in 1997 and 1993	-	.002	-0.009	.008	-.007
		(.009)	(.011)	(.010)	(.011)
ΔAge^2	.001***	.001***	.001***	.002**	.003***
	(.0001)	(.0001)	(.0001)	(.0008)	(.0009)
Experience	-.002	-.002	-.002	.00007	-.0002
	(.002)	(.002)	(.002)	(.002)	(.002)
Experience^2	.00005**	.00005**	.00005*	.00002	-.00003
	(.00002)	(.00002)	(.00003)	(.00002)	(.00003)
ΔLog State Income	-.015	-.011	-.011	.016	-.007
	(.051)	(.051)	(.055)	(.054)	(.06)
ΔMarried	.007	.007	.021**	.012	.026**
	(.009)	(.009)	(.010)	(.010)	(.011)
Age^2	-	-	-	-.00008	-.0002**
				(.00007)	(.00007)
Male	-	-	-	.053***	.056***
				(.008)	(.009)
Married	-	-	-	.018*	.014
				(.010)	(.011)
Black	-	-	-	-.037*	.002
				(.021)	(.054)
Log State Income	-	-	-	.128***	.126***
				(.034)	(.037)
Public College	-	-	-	-.038***	-.04***
				(.009)	(.010)
Normalized College GPA	-	-	-	-.001	-.001
				(.003)	(.003)
College Quality	No	No	No	Yes	Yes
SAT-ACT Quartile	No	No	No	Yes	Yes
Major Dummies	No	No	No	Yes	Yes
Parents' Education	No	No	No	Yes	Yes
Graduate Degree	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	.032	.032	0.027	0.073	0.069
Observations:	7686	7686	6329	7248	5977

Notes: See Notes to Table 3.

Table 8: Instrumental Variable Estimates
First Stage
Dependent Variable: Speaks Foreign Language in 1997
Linear Probability Model

	(1)	(2)
HS Requirement	.188***	.212***
	(.049)	(.050)
HS Elective	.072***	.074***
	(.021)	(.023)
College Requirement	.024*	.031**
	(.013)	(.014)
HS Requirement·College Req.	-.125**	-.119**
	(.054)	(.053)
HS Elective·College Req.	.004	-.043
	(.028)	(.028)
Other Controls?	No	Yes
R^2	.006	.127
Observations:	7705	7179

Notes:
1. Rhode Island, Texas, and the District of Columbia required high school students to take foreign language courses; California, New Hampshire, Oklahoma, Oregon, Virginia, and West Virginia included foreign language courses among the elective courses needed to fulfill graduation requirements in 1989. A complete list of college graduation requirements is available upon request from the authors.
2. Column 2 includes controls for age, age squared, experience and experience squared, marital status (married or not), race, the log of the average income in the state of residence, an indicator of attending a public college, normalized GPA, the quality of the college the individual attended, major, parents education level, whether the individual holds a graduate degree, and other high school requirements.
3. See Notes to Table 3 for a sample and variable description.

Table 9: Instrumental Variable Estimates
Dependent Variable: Log Hourly Wage in 1997

	(1)	(2)	(3)	(4)
Sample:	All	Spoke English at Home	Spoke English at Home, American Citizen, Parents Born in US	Spoke English at Home, Works>35h.
Speaks F.L.	.271*	.198	.214	.169
	(.147)	(.174)	(.184)	(.182)
Age	.031***	.035***	.035***	.035***
	(.008)	(.008)	(.008)	(.009)
Age^2	-.0003***	-.0003***	-.0003***	-.0003***
	(.0001)	(.001)	(.002)	(.0001)
Experience	.003	.003	.004*	-.002
	(.002)	(.002)	(.002)	(.002)
Experience^2	-.00003	-.00003	-.00004	.0003
	(.00003)	(.00003)	(.00003)	(.00003)
Male	.072***	.075***	.073***	.057***
	(.010)	(.010)	(.011)	(.010)
Married	.053***	.050***	.058***	.039***
	(.013)	(.013)	(.015)	(.013)
Black	-.037	-.042	.004	-.044
	(.059)	(.054)	(.061)	(.060)
Log State Income	.463*** (.041)	.469*** (.039)	.453*** (.040)	.408*** (.040)
Normalized College GPA	.008 (.009)	.005 (.003)	.005 (.004)	.003 (.003)
Public College	.008	.003	-.004	-.006
	(.010)	(.011)	(.011)	(.011)
Parents' Education	Yes	Yes	Yes	Yes
Major	Yes	Yes	Yes	Yes
SAT-ACT Quartile	Yes	Yes	Yes	Yes
College Quality	Yes	Yes	Yes	Yes
Graduate Degree	Yes	Yes	Yes	Yes
Other HS Req.	Yes	Yes	Yes	Yes
Observations:	7179	6775	6231	5674

Notes:

1. Instruments: State level foreign language requirement in high school in 1989, foreign language requirement among electives in high school in 1989 at the state level, indicators that college the individual attended had a foreign language requirement, and interactions of the previous dummy variables. Rhode Island, Texas and the District of Columbia required high school students to take foreign language courses; California, New Hampshire, Oklahoma, Oregon, Virginia and West Virginia included foreign language courses among the elective courses needed to fulfill graduation requirements in 1989. A complete list of college graduation requirements is available upon request from the authors.
2. Other High School Requirements is a set of variables that capture the requirements in the state where the individual studied in High School. They include English, Social Studies, Math, and Science. See Appendix Table A.5 for a complete description of these requirements.
3. See Notes to Table 3 for a sample and variable description.

Table A.1: Summary Statistics

Bachelor and Beyond Sample, 1997 Survey

Panel A: Demographics

Variable:	Mean	Standard Dev.	Min.	Max
Log Hourly Wage	2.55	.420	1.03	3.73
Speaks FL	.340	.474	0	1
Age	30.102	6.543	18	73
Experience	43.277	11.508	0	57
Married	.305	.460	0	1
Black	.040	.196	0	1
Male	.448	.497	0	1
Normalized GPA	3.063	.498	1.17	4
Public College	.677	.467	0	1
MBA	.028	.167	0	1
Ph.D.	.029	.168	0	1
Other Masters	.256	.436	0	1
Observations: 7940				

Panel B: Regional Distribution of Observations (State of Residence)

Region:	
New England	.10
Middle Atlantic	.12
East North Central	.15
West North Central	.08
South Atlantic	.20
East South Central	.06
West South Central	.11
Mountain	.06
Pacific	.12

Notes:

1. Subsample of Individuals who have hourly wages above $2.8 (1 percentile) and below $42.3 (99 percentile), who answer the question on whether they speak a foreign language and with complete data on age, experience, gender, marital status, race, state of residence, college GPA, and type of college attended.

2. Mean values weighted using sample weights.

3. New England: CT, ME, MA, NH, RI, VT; Middle Atlantic: NJ, NY, PA; East North Central: IN, IL, MI, OH, WI; West North Central: IA, KS, MN, MO, NB, ND, SD; South Atlantic: DE, DC, FL, GA, MD, NC, SC, VA, WV; East South Central: AL, KY, MS, TN; West South Central: AR, LA, OK, TX; Mountain: AZ, CO, ID, NM, MT, UT, NV, WY; Pacific: AK, CA, HI, OR, WA.

Table A.2: Percentage of Second Language Speakers and Average Earnings by Major, 1997 Survey

	% Speak a FL	Average Log(Wage)
Health	.214 (.021)	2.818 (.019)
Engineering	.216 (.022)	2.844 (.020)
Computer Sciences	.219 (.036)	2.774 (.032)
Vocational/Technical	.226 (.036)	2.524 (.032)
Business	.227 (.011)	2.599 (.010)
Education	.271 (.014)	2.384 (.012)
Life Sciences	.277 (.021)	2.404 (.019)
Other Technical/Professional	.318 (.018)	2.530 (.016)
Social Sciences	.330 (.014)	2.489 (.012)
Mathematics	.346 (.044)	2.519 (.039)
Physical Sciences	.382 (.047)	2.496 (.041)
Humanities	.450 (.017)	2.428 (.015)

Table A.3: Estimated Effect of Speaking a Foreign Language
Propensity Score Estimation
Logit Model

Dependent Variable: Speaks F.L. in 1997	
	Coefficient
Age	-.103** (.041)
Age^2	-.001** (.0005)
Male	.006 (.052)
Black	.353** (.162)
Log State Income	.650*** (.165)
Spoke English at Home	-1.848*** (.144)
Mother Born in US	-1.059*** (.102)
Mother's Education	Yes
Father's Education	Yes
Observations:	7684

Notes:

1. Observations weighted using sample weights.
2. *** statistically significant at the 1% level, ** at the 5% level, * at the 10% level.

Table A.4: Descriptive Statistics: Pre-Determined Covariates
Matched Sample

Sample:	*Speak*	*Not Speak*	*Difference*
Age	29.648	29.39	.249
	(6.142)	(5.61)	(.166)
Black	.088	.071	.017
	(.283)	(.257)	(.007)
Male	.444	.461	-.016
	(.496)	(.498)	(.014)
Log State Income	9.80	9.80	.005
	(.157)	(.148)	(.004)
Spoke English at Home	.770	.908	-.137***
	(.420)	(.288)	(.009)
Mother Born in US	.758	.870	-.111***
	(.427)	(.335)	(.010)
Observations:	3097	2451	5548

Notes:

1. Sample of individuals who have hourly wages above $2.8 (1 percentile) and below $42.3 (99 percentile), and who answer the question of whether they speak a foreign language. Each individual who speaks a foreign language was matched with an observation from the subsample of individuals who don't speak a foreign language with the closest propensity score. The matching was with replacement (i.e., each control observation was allowed to be the match for more than one *treated* observation).
2. Mean values weighted using sample weights, taking into account that the matching was with replacement.

Table A.5: High School Graduation Requirements in 1989

State:	F.L Req.	F.L. Elec.	All	English	Social Studies	Math	Science	Physical Ed.	Electives
Alabama	0	0	20	4	3	2	1	3.5	6.5
Alaska	0	0	21	4	3	2	2	1	9
Arizona	0	0	20	4	3	3	3	1	6.5
Arkansas	0	0	20	4	3	3	3	1	6.5
California	0	1	13	3	3	2.5	2.5	2	-
Colorado[1]	0	0	-	-	-	-	-	-	-
Connecticut	0	0	20	4	3	3	2	1	6
Delaware	0	0	19	4	3	2	2	1.5	6.5
D.C.[2]	1	0	20.5	4	2	2	2	1.5	8
Florida	0	0	24	4	3	3	3	1	9
Georgia	0	0	21	4	3	2	2	1	8
Hawaii	0	0	20	4	4	2	2	1.5	6
Idaho	0	0	20	4	2	2	2	1.5	6
Illinois	0	0	16	3	2	2	1	4.5	2.25
Indiana	0	0	19.5	4	2	2	2	1.5	8
Iowa[3]	0	0	-	-	1.5	-	-	1	-
Kansas	0	0	20	4	3	2	2	1	8
Kentucky	0	0	20	4	2	3	2	1	7
Louisiana	0	0	23	4	3	3	3	2	7.5
Maine	0	0	16	4	2	2	2	1.5	3.5
Maryland	0	0	20	4	3	3	2	1	5
Massachusetts[3]	0	0	-	-	1	-	-	4	-
Michigan[3]	0	0	-	-	0.5	-	-	-	-
Minnesota	0	0	20	4	3	1	1	1.5	9.5
Mississippi	0	0	16	3	2.5	1	1	0	8.5
Missouri[4]	0	0	24	4	3	3	3	1	8
Montana	0	0	20	4	1.5-2	2	1	1	10.5-10
Nebraska[5]	0	0	-	-	-	-	-	-	-

Table A.5, Continued: High School Graduation Requirements in 1989

State:	F.L Req.	F.L. Elec.	All	English	Social Studies	Math	Science	Physical Ed.	Electives
Nevada	0	0	20	3	2	2	1	2.5	9.5
New Hampshire	0	1	19.75	4	2.5	2	2	1.25	4
New Jersey	0	0	18.5	4	2	2	1	4	4
New Mexico	0	0	21	4	2	2	2	1	9
New York	0	0	18.5	4	4	2	3	0.5	varies
North Carolina	0	0	20	4	2	2	2	1	9
North Dakota	0	0	17	4	3	2	2	1	5
Ohio	0	0	18	3	2	2	1	1	9
Oklahoma[4]	0	2	15	4	2	3	2	0	4
Oregon	0	1	22	3	3.5	2	2	2	8
Pennsylvania	0	0	21	4	3	3	3	1	2
Rhode Island[6]	2	0	18	4	2	3	2	0	4
South Carolina	0	0	20	4	3	3	2	1	7
South Dakota	0	0	20	4	3	2	2	0	8
Tennessee	0	0	20	4	1.5	2	2	1.5	9
Texas	2	0	21	4	2.5	3	2	2	7
Utah	0	0	24	3	3	2	2	2	9
Vermont	0	0	15.5	4	3	3	3	1.5	0
Virginia	0	3	20	4	3	2	2	2	6
Washington	0	0	18	3	2.5	2	2	2	5.5
West Virginia	0	1	21	4	3	2	1	2	7
Wisconsin	0	0	13.5	4	3	2	2	2	0.5
Wyoming	0	0	18	1	-	-	-	-	-

Source: *Digest of Education Statistics*, 1985-1986. Office of Educational Research and Improvement, U.S. Department of Education Center for Statistics.

Notes:

[1]Local boards determine requirements. The state has constitutional prohibition against state requirements.

[2]For comprehensive diploma.

[3]Local Boards determine additional requirements.

[4]For college preparatory studies certificate.

[5]200 credit hours required, at least 80 percent in core curriculum courses. The state was conducting hearings to define core courses at the time of the survey.

[6]College bound degree.

Table A.6: Descriptive Statistics

Panel A: Comparison of Individuals Whose Parents Live in 1993 in States with and Without Foreign Language High School Graduation Requirements

Variable:	Req.	No Req.	Difference
Log State Income	9.808	9.779	.029***
	(.117)	(.170)	(.003)
Log Hourly Wage	2.571	2.544	.027**
	(.417)	(.421)	(.011)
Speaks FL	.398	.323	.074***
	(.489)	(.467)	(.012)
Age	29.7	30.224	-.524***
	(5.626)	(6.793)	(.172)
Experience	43.273	43.278	-.004
	(11.322)	(11.565)	(.304)
Married	.263	.318	-.054***
	(.440)	(.465)	(.012)
Black	.065	.032	.032***
	(.247)	(.177)	(.005)
Male	.446	.448	-.001
	(.497)	(.497)	(.013)
Normalized	3.048	3.067	-.019
College GPA	(.506)	(.496)	(.013)
Public College	.716	.665	.051***
	(.450)	(.472)	(.012)
MBA	.026	.029	-.002
	(.161)	(.169)	(.004)
Ph.D.	.034	.027	.006
	(.182)	(.164)	(.004)
Other Master's	.257	.255	.001
	(.437)	(.436)	(.011)
Observations:	1855	6085	7940

Notes: See Notes to Table 1.

Panel B.1: Comparison of Individuals by Existence of Requirement in College Attended

Variable:	Req.	No Req.	Difference
Log State Income	9.929	9.956	-.027***
	(.135)	(.129)	(.003)
Speaks F.L.	.344	.332	.012
	(.475)	(.471)	(.011)
Age	29.566	30.185	-1.61***
	(5.889)	(7.584)	(.154)
Experience	43.135	43.563	-.427
	(11.702)	(11.103)	(.273)
Married	.291	.334	-.042***
	(.454)	(.471)	(.010)
Black	.035	.050	-.015***
	(.183)	(.219)	(.004)
Male	.452	.439	.013
	(.454)	(.496)	(.011)
Normalized	3.055	3.078	-.023**
College GPA	(.493)	(.509)	(.011)
Public College	.752	.524	.227***
	(.431)	(.499)	(.010)
MBA	.032	.020	.011***
	(.178)	(.143)	(.003)
Ph.D.	.032	.023	.004**
	(.178)	(.151)	(.004)
Other Masters	.252	.263	-.010
	(.434)	(.440)	(.010)
Observations:	5211	2729	8624

Panel B.2: College Quality and Foreign Language Requirements

Category:	Req.	#Colleges	# Individuals in Sample
Most Competitive	.684	19	315
	(.477)		
Highly Competitive	.454	33	549
	(.505)		
Very Competitive	.745	102	1642
	(.437)		
Competitive	.613	251	3198
	(.490)		
Less Competitive	.5	96	940
	(.502)		
Non Competitive	.386	75	562
	(.490)		
Public	.614	324	4976
	(.487)		
Private	.464	278	2558
	(.499)		