Title
Three essays on the economics of immigration and education

Author
Suen, Karmen

Publication Date
2008

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Three Essays on the Economics of Immigration and Education

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Economics

by

Karmen Suen

Committee in charge:

Professor Julian Betts, Chair
Professor Eli Berman
Professor Julie Cullen
Professor Tomás Jiménez
Professor Chris Woodruff

2008

The dissertation of Karmen Suen is approved, and it is acceptable in quality and form for publication on microfilm:

Chair

University of California, San Diego

2008

DEDICATION

To my parents, brother, and the living Christ

EPIGRAPH

But he said to me, "My grace is sufficient for you, for my power is made perfect in weakness.".

—2 Corinthians 12:9

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This dissertation would have not been able to complete without the guidance given by my advisor, Julian Betts. Julian has been patient in leading me to the right direction when conducting research all these years. I would also like to thank the enormous amount of support and useful comments I have received from Eli Berman and Julie Cullen. It has been an honor being able to learn from all three of them. Suggestions and encouragement provided by my other two committee members, Chris Woodruff and Tomás Jiménez, are to be gratefully acknowledged as well.

Moreover, it has been great learning from and growing with my dear classmates at UCSD. In particular, I would like to express my deepest gratitude towards Hiroaki Kaido, Munechika Katayama, Young Do Kim, Jeffrey Lin, Young Joon Park, and Tatsuyoshi Okimoto for their help and companionship.

I am blessed to have Isaac Li for being there, supporting me to pursue what I have dreamt of. Furthermore, I would like to thank Clement Ko, Howard Lam, and Cherrie Ng for their prayer and care.

Thank you Jesus for being my Savior and guide at all times.

Most importantly, I could never have achieved all the good things I have encountered without the unconditional love and support of my parents, Christopher and Christine Suen.

VITA

2001	B.A., *Honors*, Economics, University of California, San Diego
2001	B.S., Chemistry, University of California, San Diego
2003	M.A., Economics, University of California, San Diego
2008	Ph.D., Economics, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Three Essays on the Economics of Immigration and Education

by

Karmen Suen
Doctor of Philosophy in Economics

University of California San Diego, 2008

Professor Julian Betts, Chair

In the first chapter of this thesis, the 1995 TIMSS eighth-grade mathematics score is used to proxy for home country education quality for U.S. immigrants. On average, a one standard deviation increase in TIMSS magnifies the marginal returns to post-migrational education by 0.83 percentage points. This pre-migrational education quality effect remains positive and significant for individuals at the 25th percentile of the conditional wage distribution. In addition, diminishing returns to post-migrational years of schooling is observed at all wage quantiles, but evidence is mixed in regards to pre-migrational years of education.

Using the 2000 Census, the second paper finds that, compared to another immigrant holding a job that requires less human-interaction, an immigrant worker who possesses knowledge in speaking a non-English language and who works in a human-interaction-intensive occupation would enjoy an average wage benefit of 4.47%. For an immigrant, other immigrants from a different home country are perceived as complements, while those from the same country of origin would be substitutes. Moreover, a one standard deviation increase in bilateral trade volume between the United States and the immigrant's country of origin is predicted to enhance the immigrant's returns to working in the Wholesale Trade industry by 3.36% on average, a pattern that is very different for immigrants whose country of origin uses English as an official language.

A positive relationship between parental involvement in reading-related activities before the student began schooling and the student's 2001 PIRLS test score is found in the third chapter. On average, having a parent who played alphabet toys, played word games, and read signs and labels out loud during the student's preschool years is predicted to carry an effect size of 0.2, holding other attributes constant. However, the effect of watching reading programs on television on this test score seems negative. Under a quantile regression framework, the effect of these parental inputs continues to be observed for students belonging to the 25th quantile of the conditional score distribution. Lastly, these academic variables are predicted to not affect an immigrant student's PIRLS score, although small sample size may be an issue.

Chapter 1

Labor Returns to Home Country Education Quality for Immigrants

1.1 Introduction

The primary focus of this paper is to test whether home country education quality affects an immigrant's labor returns to both pre- and post-migrational years of schooling. Although the Mincer (1970, 1974) model suggests that labor earnings are positively related to the number of years of education that an individual obtains, Hanushek and Luque (2002) emphasize the role that education quality plays in the payoff of human capital investment. Specifically, the relationship between quality of education and one's labor earnings should be positive[1]. In the existing literature, researchers typically use measures such as pupil-per-teacher ratio, per capita GDP classified as educational expenses, wage rate of local high school teachers, and both high school and college enrollment rates as proxies for academic quality (Betts and Lofstrom (2000); Barro and Lee (1996); Bratsberg and Terrell (2002); Woessmann (2003)). However, empirical analyses find mixed evidences for their effectiveness on achievement enhancement of students.

[1]Some U.S. studies also find evidence for a positive relationship between college quality and one's subsequent wage rate. See, for example, Brewer, Eide, and Ehrenberg (1999) and Hoxby (2001).

To measure the effectiveness of different education policies, test scores have been widely used as an outcome variable. For instance, if a reduction in class size induces a positive test score gain among students, then one concludes that smaller classes are beneficial to students. The frequent use of test score in the education literature shows its role in capturing the effect of certain policy impositions, such as an improvement on education quality, on an individual's outcome. As a result, this paper introduces a new proxy, similar to the one used in Hanushek and Kimko (2000)[2], namely, a country's average score in an international test, to represent its education quality. In particular, the eighth-grade mathematics score of the 1995 Trends in International Mathematics and Science Study (TIMSS) will be used[3]. Although TIMSS aims to analyze both the mathematical and scientific ability of young students, in this paper, only the eighth-grade mathematics outcome is used to proxy for pre-migrational[4] education quality. As shown in the literature, mathematics skill plays a significant role in both an individual's educational attainment and wages received[5].

The approach adopted in this paper differs from that of Hanushek and Kimko (2000) by the introduction of an interaction between home education quality and years of both home and U.S. schooling. A further distinction between my work and previous studies includes the performance of quantile regressions to test whether the effect of coming from a high-scoring country varies across the earnings distribution. That is, testing whether a person located at the lower 25th percentile of the conditional wage distribution would receive a dissimilar payoff compared to a person earning at the upper 75th percentile.

The third aim of this paper is to study whether diminishing returns to years of

[2]Using international test scores administered by both the International Association for the Evaluation of Educational Achievement (IEA) and the International Assessment of Educational Progress (IAEP) as a measure of a country's education quality, Hanushek and Kimko (2000) find that a one-point increase in test score increases individual earnings by 0.19%. Moreover, the magnitude of this quality effect becomes even greater for someone who finished all his schooling in his home country before arriving the United States: On average, earnings rise by 0.21 percentage points per unit of test score gain.

[3]A detailed description of TIMSS can be found in Appendix 1.7.1.

[4]Throughout this paper, *Home education* and *Pre-migrational education* will be use interchangeably.

[5]See Bishop (1992), Hanushek and Kimko (2000), and Rose and Betts (2001). For instance, Rose and Betts (2001) argue that students who completed intermediate algebra and calculus upon high school graduation tend to have a higher probability of finishing college. By the time they enter the labor market, their wage rates also tend to be greater than their counterparts', who did not study calculus before graduating high school.

schooling is present once total years of education is separated into pre- and post-migrational education. By dividing total years of education into two parts, pre- and post-migrational education are viewed as separate inputs in the human capital production function for immigrants. This concept is supported in previous work such as Betts and Lofstrom (2000) and Friedberg (2000). In these papers, the returns to pre-migrational education are consistently different from post-migrational education returns. This finding may be explained by the incompatibility between knowledge learned in one's home country and that needed in the country of destination's labor market. In this case, pre- and post-migrational education should be treated as two different inputs for immigrants.

According to Psacharopoulos (1993), returns to primary education are high in developing countries, but they tend to diminish either as the individual attains more schooling, or as a country's per capita income increases. If there exist diminishing returns to education, for an immigrant who finished certain years of schooling in his or her home country and continued his or her education after arriving the U.S., diminishing returns to U.S. education is more likely to take place. That is, since years of U.S. education obtained tend to occur in later years of an immigrant's total education, diminishing returns would more likely be associated with studies that took place in the United States. This suggestion would then be very different from what is found in Betts and Lofstrom (2000) and Friedberg (2000), which suggest that returns to years of education obtained in an immigrant's country of destination are higher than that obtained prior to immigration, since skills learned in the country of destination's school would fit the need of the local labor market better. In addition to testing which of these theories holds for immigrants included in the sample used in this paper, it will be interesting to note whether the validity of this diminishing returns to education hypothesis varies across individuals at different quantiles of the conditional wage distribution as well.

Given that the focus of this paper is on the effect of pre-migrational education quality on immigrant wages, it is natural to first consider the question of who migrates and who does not. Assuming the cost of relocation to be fixed, Borjas (1987) states that an individual's migration decision depends on the relative wage dispersion between one's country of origin and the receiving country. For a country that has higher returns to skill and also a wider wage dispersion compared to the United States, low-skilled workers are more likely to choose to emigrate to the U.S. (negative selection). On the contrary, if the country of origin's returns to skill and wage dispersion are both low, then high-skilled workers would favor migrating to the U.S. (positive selection). However, an exception to the theory is found in Chiquiar and Hanson (2005). Focusing on immigrants from Mexico, they do not find support for negative selection, even though the returns to skill and wage dispersion are both higher in Mexico than in the United States. Specifically, the direction of selection remains ambiguous due to possible variation in the cost of migration, violating the constant-cost assumption that Borjas (1987) uses in his immigrant selection model. If high-skilled individuals incur a lower cost of migration due to their ability to access additional information, then they would face a different cost compared to low-skilled workers. In this case, there exists a correlation between skill and migration costs. Consequently, even though the returns to skill are higher in Mexico than in the U.S., under heterogeneous migration costs, positive or intermediate selection could take place. In other words, findings in Chiquiar and Hanson (2005) suggest that the direction of self-selection cannot be inferred from wage dispersion alone.

Based on the above arguments, unless one has accessibility to individual country's immigration records to determine who decides to emigrate and whether migration cost can be assumed constant across immigrants with various skill levels, the direction of immigrant selection cannot be clearly identified. Therefore, regression results shown in this paper should be interpreted cautiously, for individual ability remains unobserved in all regressions and this selection bias among immigrants is likely to affect the predictive power of coefficient estimates presented. In particular, since the TIMSS variable used in this paper is a national measure, it may correlate with the standard deviation of earnings in one's country of origin, which further correlates with the self-selection bias, according to Borjas' (1987) hypothesis. To address this issue, earning variance in each of the countries included in the sample will be used to test for the degree of selection among immigrants. Results are shown in Section 1.5.3.

Regression results indicate that a one standard deviation increase in pre-migrational education quality[6] amplifies the returns to a year of U.S. education by 0.83 percentage points[7]. A joint one standard deviation jump in TIMSS score and one additional year of U.S. schooling is predicted to raise wages by 6.99%. Although TIMSS does not affect the returns to an additional year of pre-migrational education, the returns to pre-migrational education are greater than that of post-migrational education: an additional year of pre-migrational education is predicted to increase wages by 10.31%. Further, even though TIMSS does not affect the returns to an extra year of home education on average, it does positively magnify home education returns for immigrants at the lower 25th percentile of the conditional wage distribution. Based on quantile regression outputs, a one standard deviation increase in TIMSS increases the returns to an additional year of home education by 1.01 percentage points, while it strengthens the marginal effect of an additional year of U.S. education on wages by 1.18 percentage points for immigrants at the lower 25th quantile of the conditional wage distribution. For workers at upper percentiles, such effects become insignificant. Lastly, the diminishing returns to U.S. education hypothesis holds true at all quantiles, while support for diminishing returns to home country education is not found.

The paper is organized as the following. Section 1.2 describes the data set, while section 1.3 presents the main empirical framework used. Regression results and their interpretations will then be provided in section 1.4. In section 1.5, additional robustness checks are shown. Furthermore, concluding remarks and future work are noted in section 1.6.

[6] A separate set of regressions using a country's 1994 GDP per capita as an education quality measure is also performed, but the interaction term between GDP and years of education is consistently insignificant across all specifications. This finding shows that TIMSS score provides more power in explaining returns to education compared to GDP per capita, thus is used as an education quality measure in this study.

[7] According to Table 1.6, countries that scored high in TIMSS tend to be non-English speaking. As a result, a statistically significant impact of TIMSS on the returns to an additional year of U.S. education can be attributed to the fact that, for individuals from non-English speaking countries, arriving early in the U.S. is beneficial for adult English (See Bleakley and Chin (2003).). To test this possibility, an interaction term between years since migration and whether a person came from an English-speaking country is created. Since regression results remain robust upon incorporation of this term, additional estimates are not reported.

1.2 Data

The main data set used in this study is extracted from the 5% Integrated Public Use Microsample Series (IPUMS) version of the 2000 Census, which provides information on an individual's birthplace, hours worked per week, labor income, state of residence, gender, years since migration, and educational attainment. Since this paper's primary focus is on the effect of pre-migrational education quality on immigrant earnings, only working individuals[8] who were born in a non-U.S. country that participated in the 1995 TIMSS study are included in the sample. To calculate an individual's years of pre- and post-migrational education, which are not explicitly given in the Census, information are drawn from the discrete educational attainment variable. For example, through assuming a high school graduate to finish high school at age 18, subtracting one's age at arrival from age at graduation provides information on both pre- and post-migrational years of education. One of the main concerns about using the 1995 TIMSS as a proxy for pre-migrational education quality for individuals who participated in the 2000 Census is that it is very unlikely that the same group of individuals would have taken the 1995 TIMSS in their home country prior to arriving the United States. Given that this mismatch between workers and education quality becomes the most severe among those who are much older than 14 years of age (which is approximately the age of the 1995 TIMSS eighth-grade mathematics test-takers), I further limit the sample to include only individuals who are between 25 and 35 years old. Additional details on sample criteria are provided in Appendix 1.7.2.

In all analyses, the quality measure, TIMSS, is standardized to be mean-zero with

[8] I am aware of the potential sample selection problem which deals with the inability to observe wages of individuals not in the labor force. Specifically, each worker would face a two-stage decision problem where, in the first stage, he or she decides whether to participate in the labor force. Then in the second stage, his or her wages are modeled as a dependent variable affected by various factors. In this case, the error terms in the two stages could be correlated (For example, if unobserved innate ability affects both the individual's labor market participation decision and also wages received.). The literature has not reached an unanimous concensus on whether the presence of this selectivity bias would present an upward, downward, or no effect on the coefficient estimate on the returns to education. For instance, using the Finnish Census, Uusitalo (1999) finds the existence of an upward bias on returns to education if ability is being uncontrolled for; however, this upward bias could readily be offset by a downward bias resulted from the presence of endogeneity in the individual's schooling decision, resulting in an ambiguous sign for the overall bias.

a standard deviation of one across countries. In order to observe whether it affects an individual's outcome (as measured in terms of wages), interaction terms between TIMSS and pre- and post-migrational education variables are incorporated in the model.

1.3 Empirical Framework

The basic regression model is the following:

$$
\begin{aligned}
y_i \quad = \quad & \alpha + \beta_1 Homeeducation_i + \beta_2 Homeeducation_i^2 + \gamma(TIMSS * Homeeducation)_i \\
& + \zeta_1 USeducation_i + \zeta_2 USeducation_i^2 + \eta(TIMSS * USeducation)_i \\
& + X'\theta + \alpha_c + \alpha_s + \epsilon_i \qquad (1.1)
\end{aligned}
$$

In equation (1.1), the dependent variable, y_i, is natural logarithm of individual i's hourly wage rate. The set of explanatory variables listed includes years of home education ($Homeeducation_i$) and its square term, an interaction term between years of home education and TIMSS score ($(TIMSS * Homeeducation)_i$), total years of U.S. education ($USeducation_i$) and its square term, an interaction term between years of U.S. education and TIMSS score ($(TIMSS * USeducation)_i$), and X, a vector of individual characteristics. In particular, X includes person i's gender, years since migration and its squared term, as well as an indicator variable describing whether the person speaks English[9].

In order to take into account the unobservable traits of an individual's country of origin that correlates with his or her educational choice and thereby earnings pattern, a country fixed effects setup, as represented by a set of dummy variables contained in α_c, is used. This setup controls for factors that may affect an immigrant's educational attainment, such as a spillover effect measured by the average years of education attained by individuals in the immigrant's home country. In addition, I also introduce a set of

[9] The Census contains a question that asks a person to rate himself or herself as speaking English "not at all, some, well, or very well." For individuals who report themselves as speaking English "some, well, or very well," this indicator variable is assigned a value of 1.

dummy variables that represents an individual's state of residence in the United States (α_s), under the presence of possible unobservable state-specific factors that correlate with years of education an individual attains. For instance, state characteristics that attract only highly educated immigrants to reside in that particular state, thus creating immigrant clusters by education levels.

Although the same average TIMSS score is used to proxy for pre-migrational education quality for all individuals who came from the same country, my major source of variation lies in the different years of schooling that these individuals attain prior to and after migrating to the United States. Upon controlling for years since migration and other individual characteristics, there are also factors that the framework cannot account for. In particular, what determines an individual's educational attainment in his or her country of origin?

There are three possible explanations. First, a person's pre-migrational educational attainment may be affected by some random variations that he or she cannot directly control for. For instance, if the school-age entry law differs across countries, then immigrants of the same age but are from various countries may have achieved differing years of education before arriving the United States.[10] To combat this possibility, country fixed effects dummies are already incorporated into the regression model, thus this school entry law effect is being taken care of. Second, educational attainment can also strongly depend on one's family background, which is unobserved in this study. Not only will better educated parents induce their children to also attain higher levels of education, but they can provide better resources at home, which consequently encourage their children to do better in school. Third, an individual's age at arrival also affects the years of home education he or she has attained. Holding total education fixed, two immigrants' years of pre- or post-migrational education could differ due to their respective ages at arrival. This factor becomes the major source of variation in this paper, since the correlation between age at arrival and pre- and post-migrational years of education are 0.98 and -0.94, respectively.

[10] Another concern is what the legal age for teenagers to quit school is. According to Table 394 of the **Digest of Education Statistics, 2002**, age for compulsory attendance usually ends at at least 12 years of age, which is the average age of entry of individuals in my sample and therefore would not greatly affect their years of pre-migrational educational attainment.

Econometrically speaking, if family background and one's innate ability are not captured in the regression model, then the assumption that the covariance between years of education and the error term, ϵ_i, being zero is violated. As a result, the OLS coefficient estimate on returns to education is biased. If these unobservables positively affect one's education outcome, then the resulting coefficients will be overestimated.

However, my empirical strategy also utilizes a differences-in-differences approach to analyze how pre-migrational education quality affects one's returns to years of education. Focusing on the education variables (*Homeeducation* and *USeducation*), it is true that unobserved innate ability can upwardly bias their coefficient estimates (β_1, β_2, ζ_1, and ζ_2). However, if the degree of selection remains constant across countries, that is, if this upward bias persists for all individuals of high ability, regardless of which country they immigrate from, then the coefficient estimates for *(TIMSS*Homeeducation)* and *(TIMSS*USeducation)*, which are γ and η, respectively, should be less biased compared to coefficient estimates associated with *Homeeducation* and *USeducation*. As a result, the *(TIMSS*Homeeducation)* and *(TIMSS*USeducation)* terms would provide information on how pre-migrational education quality alters the returns associated with years of pre- and post-migrational schooling, regardless of the direction of bias associated with unobservables in the regression.

1.4 Regression Results

1.4.1 OLS Estimates

As shown in Table 1.1, the returns to human capital investment, whether achieved in one's country of origin or in the United States, are always positive and statistically significant. In particular, returns to an additional year of schooling obtained in an immigrant's home country are greater than that obtained in the United States. An additional year of U.S. education is predicted to raise an immigrant's wages by 6.16%, while an additional year of home country education is anticipated to increase the wage rate by 10.31% (column (3)). These findings are consistent with the hypothesis that wages are positively correlated with years of education, as supported in many other studies[11], and also that pre- and post-migrational education enter the human capital accumulation process as separate inputs for immigrants.

To test whether pre-migrational education quality affects earnings, concentrate on the (*TIMSS *Homeeducation*) and (*TIMSS*USeducation*) terms in column (3). From the coefficient of the interaction created between years of education and pre-migrational education quality, a one standard deviation increase in the TIMSS score magnifies the marginal effect of a year of U.S. education by 0.83 percentage points[12]. For an immigrant who came from a country that scored one standard deviation higher than average in TIMSS, an additional year of U.S. education is predicted to increase his wage rate by 6.99%. This supports the prediction that education quality in one's country, prior to arriving the United States, does affect one's U.S. education returns.

Moreover, F-statistics related to testing the nulls that all terms associated with pre-migrational education quality, as well as country dummies, are jointly zero show rejection to both of the hypotheses (Table 1.1). Regression results therefore support the need to incorporate pre-migrational education quality and country fixed effects dummies in the wage regression, for there are possibly unobserved variations correlated to one's country of origin that are not captured in the model.

In addition to the above findings, supporting evidence for diminishing returns to years of schooling in the United States can also be found in columns (4) through (6) of Table 1.1. The coefficient estimate for the square term of years of U.S. education is negative and significant. However, the diminishing returns to home education hypothesis fails to hold, for the coefficient of the $Homeeducation^2$ term is statistically insignificant (column (6)). It is interesting to note that, upon inclusion of the square terms, the returns to both pre- and post-migrational years of education become larger than that predicted in column (3). In particular, an extra year of home education is associated with a 10.78% increase in wages, while an additional year of U.S. education is expected to increase wages

[11] For instance, Mincer (1970, 1974).

[12] Pre-migrational educational quality does not seem to enhance the returns to an additional year of home education, for the coefficient of the (*TIMSS*Homeeducation*) term (0.0055) is statistically insignificant.

by 6.48% (column (6))[13]. Additionally, similar to the case where the square terms of pre- and post-migrational education are omitted (column (3)), the coefficient associated with the (*TIMSS*USeducation*) term is statistically significant, positive, and even greater than that obtained in column (3). When quadratic terms of education are incorporated in the model, a one standard deviation increase in pre-migrational education quality is predicted to raise the returns to a year of post-migrational education by 0.88 percentage points, which is 0.05 percentage point higher than when years of schooling is modeled as a linear function. To determine whether these findings hold true among immigrants with differing earnings patterns, I now turn to quantile regression estimate results.

1.4.2 Quantile Regression Results

To draw conclusions on whether the effect of pre-migrational education quality on immigrant wages is dependent on one's earnings, regression analysis on the 25th, 50th, and 75th percentiles of the conditional wage distribution has been performed. Results are summarized in Table 1.2[14].

Columns (1), (3), and (5) of Table 1.2 report regression estimates under the exclusion of the $Homeeducation^2$ and $USeducation^2$ regressors. Comparing workers at various quantiles, returns to an additional year of either home or U.S. education are the greatest for workers at the lower 25th percentile of the conditional wage distribution. In addition, returns to one year of education accomplished in one's home country are consistently greater than that attained in the United States at all quantiles. The differentials between home and U.S. education returns are 3.66% for the 25th, 3.69% for the median, and 3.89% for those at the 75th percentiles of the conditional wage distribution. These differentials are statistically significant (p-values of the corresponding tests are always ≤ 0.05). Additionally,

[13]In what follows, all interpretations are drawn from evaluation from the mean. For example, the returns to an extra year of home education is calculated using $\frac{\partial y}{\partial Homeeducation}=\hat{\beta}_1+(2*\hat{\beta}_2*$Average years of home education), according to Equation (1).

[14]For all regressions, I reject the hypothesis that all regressors interacting with TIMSS are jointly zero, which consequently supports the importance of incorporating pre-migrational education quality measures in the wage determination framework. In addition, the null hypothesis that all country dummy variables are jointly zero can be rejected across all quantiles, showing once again that country fixed effects should be included in the model to capture unobserved factors that are country-specific and can alter an immigrant's total educational attainment.

these results may be driven by what the optimal educational attainment accomplished by individuals at various quantiles is. When individuals at the bottom quantile of the conditional wage distribution have less years of education compared to those at the top quantile, their returns to education would also be higher compared to the latter group, since they are less likely to have reached their optimal year of education that brings them the greatest associated labor returns. As a result, these workers are also more likely to face higher returns to an additional year of education compared to those at the 75th percentile of the conditional wage distribution.

Regarding the interaction terms between home education, U.S. education, and TIMSS, they are repeatedly insignificant among upper quantiles. That is, for those at the 50th and 75th quantiles of the conditional wage distribution, pre-migrational education quality does not enhance the marginal gain associated with spending an extra year studying, regardless of whether the studying takes place in one's home country or in the United States. Contrarily, pre-migrational education quality does affect the returns to both home and U.S. education for workers at the lower 25th percentile. Precisely, for an individual at the 25th percentile of the conditional wage distribution and whose home country scored one standard deviation higher in TIMSS, marginal returns to a year of home education increases to 11.01% (compared to 10.00%), while the returns to an additional year of U.S. education becomes 7.52% instead of 6.34%.

In order to draw comparison against those obtained under exclusion of the square terms of both pre- and post-migrational years of education, results achieved under incorporation of these square terms into the original regression framework are also reported (columns (2), (4), and (6), Table 1.2). The major difference between modeling the returns to education as a quadratic instead of a linear function is how pre-migrational education quality affects the individuals at the median percentile of the conditional wage distribution. When education is modeled as a linear function, TIMSS does not affect education returns for those at the 50th percentile; upon inclusion of the quadratic education terms, TIMSS is predicted to have a positive impact on the returns to a year of post-migrational education for individuals at the median percentile (column (4), Table 1.2). Specifically, TIMSS increases the returns to an extra year of post-migrational education by 0.50 percentage points when the returns to education is modeled as a parabolic function. Although pre-migrational education quality carries a positive impact on the returns to education for individuals at the 25th and the 50th percentiles, similar to results obtained under omission of the squared years of education terms (column (5)), it has no effect on the education returns for those at the 75th percentile of the conditional wage distribution[15].

Based on quantile regression results, the diminishing returns to U.S. education presumption finds support across all quantiles. That is, the coefficient estimates of the $USeducation^2$ term are consistently negative and significant. However, signs associated with the estimates for the squared home education term change across quantiles. In particular, only at the 75th percentile is the $Homeeducation^2$ estimate statistically significant. Since the coefficient estimate is positive instead of negative (column (6)), years of pre-migrational education is expected to bring increasing returns for those at the 75th quantile of the conditional wage distribution.

So far, the effect of pre-migrational education quality on the returns to an additional year of schooling seems to be affected by whether education is modeled as a linear or quadratic function in the regression framework. As a form of robusness check for model specification, as well as testing whether the increasing returns to education for individuals at the 75th percentile of the conditional wage distribution is dependent on how education is projected, a regression framework that models education semi-parametrically is adopted in the next section.

1.5 Further Tests: Discrete Education Years

1.5.1 Discrete Years of Education Results

The main regression framework adopted so far imposes a quadratic form of years of education on immigrants. However, results presented in the previous section show a

[15]Similar to OLS results obtained upon adding square terms of years of education (compare columns (3) and (6) of Table 1.1), the returns to post-migrational schooling are magnified once $Homeeducation^2$ and $USeducation^2$ are included in the model. For example, among individuals at the 25th percentile, an additional year of post-migrational education is predicted to increase wages by 6.79% (column (2)), as opposed to the 6.34% predicted in column (1).

possible increasing returns to scale with respect to years of pre-migrational education for individuals at the 75th percentile of the conditional wage distribution, which is counterintuitive. As a result, performing regressions free of this constraint on functional form is what will be accomplished next. To achieve this goal, I create a set of dummy variables representing the specific year of pre- and post-migrational education that each individual has achieved. That is, a new set of dummy variables is defined to denote each year of pre- and post-migrational education attained by individual i[16]. In all of the regression results presented below, individuals who have zero to eight years of home and/or U.S. education are classified as the comparison group.

Under this regression framework, all of the dummy variables signifying grades 11 or above are statistically significant, regardless of whether the study has taken place in one's home country or in the United States (Table 1.3)[17]. For instance, compared to a person who has less than an 8th grade education at home and in the United States, a home college graduate is expected to earn 116.89% higher in terms of wages, while a U.S. college graduate is expected to earn 36.75% more (column (1))[18]. Judging from this result, the returns associated with a home college degree is greater than that associated with a U.S.

[16]In particular, dummy variables are classified according to information given in the Census. For a person who has studied 9 years in his or her home country prior to arriving the United States, *Homegrade9* receives a value of one, signifying the pre-migrational grade level an individual has achieved. More specifically, *Homegrade9* will be set to one for anyone who has studied 9 or more years in his or her home country. Analogously, *USgrade9* is assigned a value of one for an individual who has attained at least a 9th-grade education in the United States. To demonstrate how these dummy variables are set, consider an immigrant who finished high school in his or her home country, then became a college graduate in the United States. In this case, *Homegrade9*, *Homegrade10*, *Homegrade11*, *Homegrade12*, *USgrades13to15*, and *USgrade16* are all assigned a value of one, while *Homegrades13to15*, *Homegrade16*, *USgrade9*, *USgrade10*, *USgrade11*, and *USgrade12* are set to equal zero. Note that I have not decomposed the "grades 13 to 15" group further, since that is the category originally given in the Census.

[17]One of the puzzling results observed from Table 1.3 is the negative and statistically significant coefficient of the *USgrade12* term. This coefficient estimate predicts that, compared to those with less than 8 years of home and/or U.S. education, a U.S. high school graduate would earn 47.18% less on average. Such counterintuitive result could be due to the high collinearity between the *USgrade11* and *USgrade12* variables (The correlation between them is 0.90.). Since an individual with 12 years of U.S. education is assigned a value of one to both the *USgrade11* and *USgrade12* terms and the difference between the number of observations in these two categories is relatively small, it may be the high correlation between *USgrade11* and *USgrade12* that drives the coefficient estimate of *USgrade12* to become negative. Upon omission of *USgrade11* from the regression framework (not shown), the coefficient estimate for *USgrade12* no longer carries the counterintuitive negative sign.

[18]The hypothesis that earnings associated with a home and U.S. college degree are identical is rejected, with a p-value of 0.0000.

college education. Given that I have not "corrected" the selection bias problem on who chooses to migrate to the United States, two interpretations could be drawn. If ability is perceived as a pure signal (i.e., employers view education as just an indication of how dedicated the person is), then, from the employer's perspective, immigrants who finished their college education prior to arriving the United States could be viewed as being more able compared to other immigrants who did not finish college in their home country. As a result, the returns to having a home college degree would be greater than that associated with a U.S. college degree. On the other hand, suppose ability is interpreted as a pure human capital indicator. In this case, a positive differential between the returns associated with a home and a U.S. college degree suggests that home college education adds more value to an immigrant's human capital than a U.S. college education. Combining this finding with that observed from the quantile regression results, it is possible that low-skilled immigrants are the ones who choose to migrate to the United States. In Table 1.2, the returns to an extra year of home education are always remarkably higher than U.S. education for those at the 25th percentile of the conditional wage distribution. As discussed before, this may be due to this specific group of workers having not attained their optimal year of education, therefore incurring positive returns to pre-migrational human capital investment. In Table 1.3, the payoff associated with a home college degree is also higher compared to that of a U.S. college degree. If it is the low-skilled immigrants, who have not attained their optimal year of schooling, who choose to migrate, then having a college degree prior to migration should carry a huge benefit for them, for it distinguishes them from their lower-skilled counterparts. The validity of this suggestion of negative selection will be further explored in the last subsection.

Focusing on the interaction terms between TIMSS and each respective grade level, one notices that pre-migrational education quality is predicted to only magnify the returns associated with a U.S. college degree. Specifically, a one standard deviation increase in TIMSS enhances the returns to a U.S. college degree by 4.91 percentage points. Since there exists a positive differential between a home and a U.S. college degree, this finding suggests that for immigrants who did not finish college prior to migration, attaining a U.S. college degree and coming from countries with a better education system pay off, although it remains difficult to earn a comparable wage rate with those who finished college in their country of origin[19].

1.5.2 Discrete Years of Education with (Country of Origin × Age at Arrival) Fixed Effects Results

In the previous section, grade equivalent dummies have been created to signify the grade level attainment in both the home country and the United States for an immigrant. The use of such framework could induce another complication. For an individual who has attained a U.S. 11th-grade education, he or she could have studied up until 10th grade in the respective home country before immigrating to the United States, or he or she could have obtained all education, since 1st- to 11th-grade, solely in the United States. In this case, a U.S. grade dummy variable would not distinguish between the two possible venues, although the returns to an 11th-grade education in the U.S. could reflect the contribution of studying more in one's country of origin. More specifically, since there is no interaction term created between *USgrade11* and various home education variables, it is possible that the coefficient estimate of *USgrade11* also partially captures the accumulated effect of completing various years of home education on the returns to completing an 11th-grade in the U.S. To solve this complication, the fixed effects framework adopted so far will be altered. In particular, instead of just including dummy variables for each country of origin, interaction terms between country of origin and age at arrival will be introduced. By doing so, regression results presented take into account the individual's age at arrival, which reflects years of home education he or she has gotten, as well as the country that he or she

[19] To study how the returns to an additional year of education differs for individuals coming from countries that score high (or low) in TIMSS, another set of regression, which divides countries into high and low TIMSS-scoring, is run. Using the fact that the TIMSS variable is standardized to have a zero mean, countries are classified as "high TIMSS-scoring" if their national TIMSS score is greater than zero. Likewise, "low TIMSS-scoring" countries are those that score less than zero in TIMSS. A one standard deviation increase in TIMSS is predicted to magnify the returns associated with a U.S. 10th-grade degree for immigrants from a high-TIMSS, and that of a U.S. high school degree for those originating from a low-TIMSS country, respectively. Since most of the results remain similar to that reported in Table 1.3, they are not repeated here. Also, a separate set of regression, which includes only male workers in the sample, is run. This is performed in order to investigate whether regression results are driven by the inclusion of female workers, who tend to have a more elastic labor supply. Again, since similar results are obtained, they are not reported separately.

came from. Notice that the use of this framework will no longer permit the incorporation of pre- and post-migrational education as separate inputs for the individual. As a result, all of the education variables shown in Table 1.4 represent total educational attainment.

Using individuals with less than an elementary degree as the comparison group, those who have at least some college education are expected to earn a statistically significant 33 to 34% more, depending on whether pre-migrational education quality is incorporated in the regression model. These returns to human capital investment is further magnified for those with a college degree. On average, a college graduate is expected to earn 65 to 66% more than someone with less than an elementary degree. Contrasted with what has been found so far, pre-migrational education quality does not affect the returns to education. None of the TIMSS variables are statistically different from zero, although one does reject the hypothesis that all TIMSS variables are jointly zero (p-value=0.0076). This finding can be explained by the construction of the (TIMSS*education) and the (age at arrival*country of origin) interaction variables. In particular, since TIMSS is a national measure that remains identical for immigrants who came from the same country, its effect on wages could have been captured in the set of country fixed effects dummies. Likewise, due to the high correlation between years of pre-migrational education and age at arrival, inclusion of age at arrival in the regression framework would lessen the effect of education on immigrant wages. As a result, incorporating (age at arrival*country of origin) into the model could drive the coefficient estimates associated with (TIMSS*education) to become statistically insignificant. Unless TIMSS can be measured on an individual basis, it remains difficult to distinguish between the effects of (TIMSS*education) and (age at arrival*country of origin) on wages for immigrants.

In addition, as evident in Table 1.3, among the twelve (TIMSS*education) interaction terms, only the (*TIMSS*USgrade16*) coefficient is statistically significant. Upon using an (age at arrival*country of origin) framework (Table 1.4), the effects of a U.S. grade 16 and a home grade 16 on wages are forced to be identical. Since the coefficient estimate on (*TIMSS*Homegrade16*) is statistically insignificant with a relatively large standard error (column (2) of Table 1.3), it may be drive the coefficient on (*TIMSS*College Graduate*) in Table 1.4 to appear statistically insignificant, for it combines the effect of the positive

(*TIMSS*USgrade16*) and the negative (*TIMSS*Homegrade16*) on wages, resulting in a possible negative and statistically insignificant estimate.

Another interesting observation is drawn with respect to the fixed effects framework utilized in Table 1.4. Using the country of origin interacted with age at arrival set of dummies, one fails to reject the hypothesis of them being jointly zero. Although this revised set of fixed effects variables help to capture the effect of age at arrival and country of origin on an individual's wage rate, doing so makes it impossible to decompose an individual's educational attainment into pre- and post-migrational years of schooling. This could result in a further complication, for other studies have shown the returns to home and U.S. education differ for immigrants. Since the main goal of this paper is to test whether pre-migrational education quality affects an immigrant's returns to an additional year of both home and U.S. education and inclusion of the (age at arrival*country of origin) fixed effects framework predicts an ambiguous interpretation on the (TIMSS*education) coefficients, results obtained under the original country fixed effects model may be more useful in drawing implications (Tables 1.1 through 1.3).

1.5.3 Using Income Share of Each Country to Test the Degree of Selection on Migration Decision

The largest complication associated with this study is the degree of selection among immigrants. As proposed in Borjas' (1987) paper, positive selection is more likely to take place if the wage dispersion and the returns to skill are both low in one's country of origin, while negative selection tends to occur when they are high, compared to that of the country of destination's. In this section, an inequality measure derived by Deininger and Squire (1996) is used to investigate the possibility of signing the direction of selection among immigrants in the sample used in this paper.

One of the inequality measures provided in Deininger and Squire (1996) is the share of income earned by individuals belonging to various quintiles of the income distribution. Specifically, the ratio of income shares between individuals at the top and the bottom quintiles[20]. Upon running a correlation matrix between variables, the correlation between this inequality measure and wages is found to be negative (-0.03). What this means is that, among immigrants included in the sample, the more unequal income is in one's home country, the lower the wages would one earn[21] . In addition, with the correlation between TIMSS and *inequality* being 0.71, countries with a wider income dispersion seem to be associated with a higher education quality. Under Borjas' (1987) hypothesis, negative selection takes place when the returns to education and income dispersion are high in one's country of origin. If pre-migrational education quality positively correlates with the returns to education in the local labor market, then more able individuals would have a lower incentive to leave their country to migrate to another country. It is, therefore, possible that immigrants included in the data set negatively select themselves to come to the United States, while their high ability counterparts choose to stay in their home country.

Chiquiar and Hanson (2005) reconcile their finding of intermediate or positive selection among Mexican immigrants to a hetergeneous migration cost that is further variable depending on the emigrant's ability, while Borjas' (1987) model assumes cost to be homogeneous across immigrants. Using the distance between the capital city of the United States (Washington, D.C.) and the capitals of various immigrants' countries of origin as a proxy for the cost of migration to study the relationship between migration cost and wages, the correlation between these two variables is found to be -0.02. In other words, for an immigrant who came from a country that has a greater distance from the United States, his or her wages in the U.S. labor market would be lower on average. With a correlation measure of 0.74 between *TIMSS* and *distance*, countries that are farther away from the United States seem to be associated with a higher education quality. In other words, assuming *distance* captures migration cost, immigrants who face a higher cost of

[20] Since income share data for Cyprus and Iran are not available, regression results presented in this subsection do not include immigrants from these two countries in the sample.

[21] Upon inclusion of this inequality measure in the framework that models education semi-parametrically, no significant deviation on the (*TIMSS*education*) coefficient estimates is found. Namely, a one standard deviation increase in pre-migrational education quality is predicted to only affect the returns to attaining a U.S. college degree. However, the magnitude associated with this estimate is found to be 0.0555, which is greater than the 0.0491 reported in Table 1.3. Since no other major change is observed, results are not reported.

migration are associated with a higher pre-migrational education quality. This argument is in accord with Chiquiar and Hanson's (2005) explanation that migration cost tends to positively correlate with ability. Highly able individuals may face a lower migration cost because they are better at collecting information on migration possibilities.

The above observations suggest the sign of immigrant selection might be negative. In the sample used in this paper, it is the less able immigrants who are more likely to choose to come to the United States. If this is the case, then the value of attaining a college degree prior to immigration should be high among immigrants, which serves as an explanation to why the returns associated with a home college degree are relatively larger than that associated with a U.S. college degree (Table 1.3). Specifically, to a U.S. employer who possesses incomplete information on an immigrant applicant, negative selection would be perceived to be more likely to take place if the applicant came from a country with a high income inequality. However, if this applicant has obtained a high educational attainment, such as completion of a college degree before moving to the United States, then this degree serves as a signal that such applicant is of high ability, which distinguishes him or her from other applicants originated from the same country. As a result, the returns to holding a pre-migrational college degree would be higher than that associated with a U.S. college degree.

1.6 Conclusion

Using the 1995 Trends in International Mathematics and Science Study (TIMSS) as a means to capture pre-migrational education quality, this paper found that pre-migrational education quality is predicted to magnify the returns to an additional year of post-migrational education. A one standard deviation increase in TIMSS increases the returns to a year of U.S. education by a magnitude ranging from 0.50 (column (4), Table 1.2) to 4.91 percentage points (column (2), Table 1.3). This finding proposes the complementarity between home education quality and U.S. education returns. Specifically, a better education system in one's home country sharpens the returns associated with an additional year of post-migrational education. Moreover, there consistently exists a positive differential between home and U.S. education returns, regardless of whether the education variables are defined as continuous or as discrete dummies. Such finding suggests that, for immigrants, pre- and post-migrational years of education should be treated as separate inputs in the human capital production function.

What else do these results suggest? First, that home country education quality matters. In addition to devoting resources that encourage individuals to attain higher education, policy makers should pay equal attention to the quality of schooling, since it affects future labor outcomes. An investment in education quality widens the spectrum of skills for an individual who leaves one's country to migrate to another country, which can subsequently be treated as inducing a positive externality associated with the education quality of one's country of origin. From an efficiency point of view, to maximize worker productivity, it may be beneficial for the U.S. immigration policy to encourage more immigrant inflow from countries with a higher than average education quality. By doing so, workers who are the most productive are attracted to come, thereby increases the competitiveness of the U.S. labor force. Second, that education is an accumulative process. That is, for immigrants, the quality of education that they have received in their home country at time *t-1* tends to also affect their labor returns to education attained at time *t*. A possible extension to this argument suggests the importance of controlling for previous education quality in order to completely capture the effect of total years of schooling on wages. Not only is controlling for pre-migrational education quality essential, but it may be beneficial for policy makers to focus on how to allocate resources for immigrants from countries with a lower education quality in order to ensure the workers' competitiveness compared to those from countries with better education resources.

It is equally important to take note of some possible problems associated with the findings of this paper. Specifically, since the direction of immigrant selection seems to be negative, results obtained may be driven by immigrants who self-select into moving to the United States. With the correlation between TIMSS and the income inequality measure used being 0.71, for the countries included in the sample, a higher income dispersion implies a higher education quality. If education quality further correlates with the returns to education in an immigrant's country of origin, then high ability individuals would have a smaller incentive to migrate to the United States. As a result, less skilled individuals would choose to leave their home country, driving a negative selection process that complies with Borjas' (1987) theory.

In addition, due to data limitation, the pre- and post-migrational years of education are constructed under the use of various assumptions. As a result, education variables may not fully capture the true years of education an immigrant has obtained in his or her home country and in the United States. Although this inaccuracy may be present, overall, results do not contradict with what has been found in the literature. That is, education quality and quantity both matter to a worker's wage rate. However, no family background variables have been controlled for in the regressions ran. As discussed previously, family background may affect an individual's total educational attainment. Since the Census does not contain specific parental education variables, this caveat has to be combatted with another data set that provides the relevant measures.

One of the future work topics includes a further investigation on whether an immigrant's educational attainment decision depends on the quality of education in his or her home country. In particular, by defining an education mismatch as the discrepancy between knowledge attained through pre-migrational education and skills demanded in the United States, would pre-migrational education quality help explain dropout behavior among immigrants? If controlling for education quality succeeds in predicting school dropout behavior, then perhaps extra resources should be devoted to immigrants who are from countries with a lower-than-average school quality. Moreover, the direction of immigrant selection deserves further investigation. One possibility of doing so is to incorporate a non-constant migration cost into Borjas' (1987) model to observe whether this changes the positive and negative selection hypotheses, since results do suggest that immigrants of different ability and who came from various countries face a dissimilar cost of migration.

Table 1.1: Returns to Home Education Quality: OLS Estimates for Workers Aged Between 25 and 35

	(1)	(2)	(3)	(4)	(5)	(6)
Years of Home Education	0.0634*	0.0651*	0.1031*	0.0978*	0.1001*	0.1064*
	(0.0027)	(0.0027)	(0.0035)	(0.0057)	(0.0057)	(0.0063)
Years of Home Education, squared				-0.0017*	-0.0017*	0.0001
				(0.0003)	(0.0003)	(0.0004)
Years of U.S. Education	0.0600*	0.0616*	0.0616*	0.0854*	0.0878*	0.0829*
	(0.0023)	(0.0023)	(0.0025)	(0.0055)	(0.0054)	(0.0055)
Years of U.S. Education, squared				-0.0008*	-0.0008*	-0.0009*
				(0.0002)	(0.0002)	(0.0002)
TIMSS		-0.2057*			-0.2141*	
		(0.0462)			(0.0462)	
TIMSS*(Home Education)		0.0080*	0.0055		0.0083*	0.0055
		(0.0030)	(0.0030)		(0.0030)	(0.0030)
TIMSS*(U.S. Education)		0.0103*	0.0083*		0.0110*	0.0088*
		(0.0026)	(0.0026)		(0.0026)	(0.0026)
F(All TIMSS coefficients are zero)		22.75	9.06		24.15	10.76
p-value		[0.0000]	[0.0001]		[0.0000]	[0.0000]
F(All country fixed effects dummies are zero)			4.98			4.94
p-value			[0.0000]			[0.0000]
Country Fixed Effects	No	No	Yes	No	No	Yes
State Fixed Effects	No	No	Yes	No	No	Yes
Demographic Controls	No	No	Yes	No	No	Yes
N	12153	12153	12153	12153	12153	12153

[1] Regression results are based on foreign-born immigrant workers aged 25 to 35 who came from a country that participated in the 1995 eighth-grade mathematics test of TIMSS, excluding those from Slovenia and the Slovak Republic due to limitation on data points available. Other regressors include English ability, gender, years since migration (and its squared term), and country and state of residence fixed effects dummies, when appropriate.

[2] * signifies 5% statistical significance. Robust standard errors shown in parentheses.

Table 1.2: Returns to Home Country Education Quality: Quantile Regression Estimates for Workers Aged Between 25 and 35

	25th Percentile		50th Percentile		75th Percentile	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Home Education	0.1000*	0.1188*	0.0959*	0.0933*	0.0962*	0.0876*
	(0.0044)	(0.0079)	(0.0029)	(0.0057)	(0.0037)	(0.0064)
Years of Home Education, squared		-0.0008		0.0006		0.0010*
		(0.0040)		(0.0030)		(0.0004)
Years of U.S. Education	0.0634*	0.0860*	0.0590*	0.0861*	0.0573*	0.0832*
	(0.0032)	(0.0070)	(0.0021)	(0.0050)	(0.0027)	(0.0055)
Years of U.S. Education, squared		-0.0009*		-0.0011*		-0.0011*
		(0.0003)		(0.0002)		(0.0002)
TIMSS*(Home Education)	0.0101*	0.0094*	0.0003	0.0008	0.0003	-0.0005
	(0.0035)	(0.0036)	(0.0024)	(0.0026)	(0.0031)	(0.0029)
TIMSS*(U.S. Education)	0.0118*	0.0129*	0.0034	0.0050*	0.0037	0.0032
	(0.0030)	(0.0030)	(0.0021)	(0.0022)	(0.0026)	(0.0025)
F(All TIMSS coefficients are zero)	8.97	13.40	6.98	11.35	5.00	6.00
p-value	[0.0001]	[0.0000]	[0.0009]	[0.0000]	[0.0067]	[0.0025]
F(All country fixed effects dummies are zero)	6.67	6.37	7.43	6.68	2.86	3.06
p-value	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	12153	12153	12153	12153	12153	12153

[1] * signifies 5% statistical significance. Robust standard errors shown in parentheses.

Table 1.3: Returns to Home Country Education Quality: Regression Results for Dividing Years of Education into Grade Levels

	(1)	(2)
Home Grade 9	0.3448	0.2779
	(0.1788)	(0.1987)
Home Grade 10	0.3785*	0.3085*
	(0.1401)	(0.1471)
Home Grade 11	1.1145*	1.0484*
	(0.1901)	(0.1714)
Home Grade 12	0.5551*	0.5193*
	(0.0984)	(0.1073)
Home Grades 13 through 15	0.6480*	0.6088*
	(0.0968)	(0.1058)
Home Grade 16	1.1689*	1.1297*
	(0.0984)	(0.1071)
U.S. Grade 9	0.1854	0.1212
	(0.1774)	(0.1972)
U.S. Grade 10	-0.0203	-0.0258
	(0.1821)	(0.1969)
U.S. Grade 11	0.6599*	0.6662*
	(0.1972)	(0.1741)
U.S. Grade 12	-0.4718*	-0.4438*
	(0.1700)	(0.1422)
U.S. Grades 13 through 15	0.0972*	0.0992*
	(0.0191)	(0.0193)
U.S. Grade 16	0.3675*	0.3670*
	(0.0146)	(0.0146)

Table 1.3: Returns to Home Country Education Quality: Regression Results for Dividing Years of Education into Grade Levels, continued

	(1)	(2)
TIMSS*(Home Grade 9)		-0.2196
		(0.2219)
TIMSS*(Home Grade 10)		-0.3286
		(0.1906)
TIMSS*(Home Grade 11)		-0.2737
		(0.1727)
TIMSS*(Home Grade 12)		-0.1589
		(0.1366)
TIMSS*(Home Grades 13 through 15)		-0.1271
		(0.1351)
TIMSS*(Home Grade 16)		-0.0970
		(0.1346)
TIMSS*(U.S. Grade 9)		-0.2165
		(0.2223)
TIMSS*(U.S. Grade 10)		-0.0634
		(0.2232)
TIMSS*(U.S. Grade 11)		0.0676
		(0.1780)
TIMSS*(U.S. Grade 12)		0.0997
		(0.1176)
TIMSS*(U.S. Grades 13 through 15)		-0.0025
		(0.0213)
TIMSS*(U.S. Grade 16)		0.0491*
		(0.0148)
F(All TIMSS coefficients are zero)		2.05
p-value		[0.0171]
F(All country fixed effects dummies are zero)	4.33	4.05
p-value	[0.0000]	[0.0000]
Country Fixed Effects	Yes	Yes
State Fixed Effects	Yes	Yes
Demographic Controls	Yes	Yes
N	12153	12153

[1] * signifies 5% statistical significance. Robust standard errors shown in parentheses.

Table 1.4: Returns to Home Country Education Quality using a (Country of Origin × Age at Arrival) Fixed Effects Framework

	(1)	(2)
Elementary Graduate	0.0195	-0.0341
	(0.2146)	(0.2349)
High School Dropout	-0.0537	-0.0517
	(0.1650)	(0.1737)
High School Graduate	0.2074	0.1923
	(0.1473)	(0.1550)
College Dropout	0.3403*	0.3295*
	(0.1469)	(0.1545)
College Gradaute	0.6645*	0.6537*
	(0.1469)	(0.1545)
TIMSS*(Elementary Graduate)		-0.1716
		(0.2445)
TIMSS*(High School Dropout)		0.0092
		(0.2054)
TIMSS*(High School Graduate)		-0.0626
		(0.1676)
TIMSS*(College Dropout)		-0.0573
		(0.1670)
TIMSS*(College Graduate)		-0.0045
		(0.1669)
F(All TIMSS coefficients are zero)		3.15
p-value		[0.0076]
F(All (Country × Age at Arrival) dummies are zero))	1.29	1.57
p-value	[0.1796]	[0.0545]
(Country × Age at Arrival) Fixed Effects	Yes	Yes
State Fixed Effects	Yes	Yes
Demographic Controls	Yes	Yes
N	12153	12153

[1] All years of education classifications shown are expressed as total educational attainment, instead of dividing them further into pre- and post-migrational years of education.

[2] * signifies 5% statistical significance. Robust standard errors shown in parentheses.

Table 1.5: Summary Statistics

Variable	Observations	Mean	Standard Deviation
ln(wage)	12153	2.75	0.73
Age	12153	30.03	3.12
Female	12153	0.47	0.50
Years in U.S.	12153	17.57	8.26
Years in U.S., squared	12153	376.80	301.96
Total years of education	12153	20.97	1.50
Age at arrival	12153	12.46	7.27
TIMSS	12153	0	1
Home education	12153	7.07	6.00
U.S. education	12153	10.07	6.88
Inequality†	11228	0	1
Distance†	11228	5165.09	2656.80

[1] Sample summary statistics drawn from the Census 2000 data, including only individuals who migrated from a country that participated in the 1995 eighth-grade mathematics test of TIMSS and met criteria described in text. Data sources: 5% IPUMS version of Census 2000; 1995 TIMSS; Deininger and Squire (1996); *http://www.indo.com*.

[2] Variables accompanied by a † signify that the sample further excludes immigrants from Cyprus and Iran, since no inequality measure is available for these countries.

[3] *Inequality* refers to the weighted ratio of income share between individuals at the top and the bottom quintiles in the immigrant's respective country of origin.

[4] *Distance* refers to the distance between the United States' capital (Washington, D.C.) and the capital of the immigrant's respective country of origin.

Table 1.6: Average Scores of Countries that Participated in the Eighth-Grade Mathematics Test of TIMSS, 1995 and 1999

Country	1995	1999
Australia	519	525
Belgium(Flemish)	550	558
Bulgaria	527	511
Canada	521	531
Cyprus	468	476
Czech Republic	546	520
England	498	496
Hong Kong	569	582
Hungary	527	532
Islamic Republic of Iran	418	422
Italy	491	485
Japan	581	579
Korea	581	587
Latvia	488	505
Lithuania	472	482
Netherlands	541	540
New Zealand	501	491
Romania	482	472
Russian Federation	535	526
Singapore	643	604
Slovak Republic‡	534	534
Slovenia‡	531	530
United States‡	492	502
Number of Countries	23	23
Average	519	521

[1] For comparison purposes, I only include countries that participated in both the 1995 and 1999 eighth-grade TIMSS mathematics test. In this table, 1995 scores are adjusted so that comparisons between a country's 1995 and 1999 performance can be drawn. Data source: Exhibit 1.3 of ***TIMSS 1999 International Mathematics Report****: Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade* (2000).

[2] Countries excluded from the sample are denoted with a ‡.

1.7 Appendix

1.7.1 TIMSS

Developed by the International Association for the Evaluation of Educational Achievement (IEA) and supported by the National Center for Education Statistics (NCES) and National Science Foundation (NSF), the purpose of TIMSS (Trends in International Mathematics and Science Study) is to compare the mathematical and scientific knowledge of school-aged children living in different parts of the world. Since 1995, tests specialized in mathematics and science have been given to fourth- and eighth-graders across countries every four years[22]. As of 2003, three test cohorts have been administered. In 1995, 42 countries participated in the eighth-grade math test. The number of participants became 38 in 1999 and increased to 46 in 2003. Within each country, 50 schools are randomly drawn to participate in TIMSS. The five main areas of focus in the mathematics test are number, algebra, geometry, data, and measurement. Sample questions can be found on *http://nces.ed.gov/nceskids/eyk/index.asp?flash=true*. In this paper, I follow results obtained from Exhibit 1.3 of ***TIMSS 1999 International Mathematics Report:*** *Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*, which reports the average scores of countries that participated both in the 1995 and 1999 TIMSS[23]. Scores for both the 1995 and 1999 TIMSS eighth-grade test of mathematics are reported in Table 1.6. There are 20 countries included in the sample, namely, Australia, Flemish-Belgium, Bulgaria, Canada, Cyprus, Czech Republic, England, Hong Kong SAR, Hungary, Islamic Republic of Iran, Italy, Japan, Republic of Korea, Latvia, Lithuania, Netherlands, New Zealand, Romania, Russian Federation, and Singapore.

Although it is plausible to use the fourth-grade mathematics scores to capture home country education quality, I choose to use the eighth-grade scores for the following reason. Among working individuals who participated in the 2000 Census, it is extremely unlikely for them to have participated in the fourth-grade TIMSS test in 1995. I understand that

[22] In 1995, tests were also given to students who were about to graduate high school.
[23] To draw valid comparisons, the 1995 scores were rescaled.

it may also be unlikely for the same group of individuals to have taken part in the eighth-grade test, but the latter test would proxy for their education background better than the former one, since the age difference between the eighth-grade test administration and individuals in the sample is less than that found under the use of the fourth-grade test. Also, the eighth-grade test scores were used because there were less participants in the fourth-grade test in 1995.

1.7.2 Sample Description

My sample consists of foreign-born working individuals who are between 25 and 35 years of age. Individuals who report themselves as still being in school are excluded. Moreover, those who came from a country that did not participate in the 1995 TIMSS eighth-grade mathematics test are omitted[24]. In order to use the natural logarithm of hourly wage rate as the outcome variable, information on total wage and salary income and weeks worked (converted into unit of hours) in the previous year are used. Workers who are either (i) unemployed; (ii) not in the labor force; (iii) report themselves as employed but have a zero wage income are further excluded from the sample. To minimize the effect of outliers, workers whose wages fall into either the bottom or top 1% of the wage distribution are dropped from the sample.

1.7.3 Derivation of pre- and post-migrational years of education

Since the Census 2000 data does not directly address the amount of education an immigrant has received in his or her home country prior to arriving the United States, the pre-migration variable, *Homeeducation*, is calculated as follows. I first subtract an individual's age from 2000 (the Census year) in order to obtain his or her year of birth. Then the year of migration variable, recorded in the Census, is utilized to find out how long that person has stayed in the United States. Subtracting birth year from this term, I deduce an age at arrival for each observation. Here, assuming education starts at age

[24]Immigrants from Slovenia and the Slovak Republic are not included in the sample due to their limited number of immigrant inflow to the United States.

six[25], one can derive the home education variable by subtracting 6 from an immigrant's age at arrival. However, this approach is problematic and can be easily illustrated with the following example. An individual who arrived the United States at age 50 does not mean that he or she has received 44 years of education at home (50-6=44). Consequently, in order to derive the appropriate measure for years of home education, I must take one's total years of schooling into consideration. Since total education and age at arrival are not measured in the same unit, I convert the former into its corresponding expected age of completion. For example, I expect an individual who has completed college to have graduated at age 22 and another who has completed high school to have graduated at age 18. However, there is a caveat associated with the education measure provided in the Census. For those who drop out from college, the Census does not specify when the person has dropped out. In this case, I take the median year, which is between completion of freshman and junior studies, and assign an age 20 to such individual. Consider the following formulation:

$$Home\ education = Total\ education - Age\ at\ arrival - 6 \qquad (1.2)$$

For an individual whose age at arrival is greater than his or her total education correspondence age, I assume the person has completed all education in the respective home country. On the other hand, for an individual arriving before age 6, I assume he or she has not received any education in his or her country of origin. For all others, years of home education is found by applying the formula above.

To calculate a person's years of U.S. education, again approximate the age of completion for each education level (For example, assuming an individual completes junior high education at age 14, high school education at age 18, and college education at age 22.), match this with the person's total educational attainment, and subtract it by years of home education to obtain the total years of education in the United States. This strategy clearly requires the following assumptions. First, that education starts at age six. Second, that

[25] According to the NCES's **Digest of Education Statistics 2002**, most of the countries' compulsory schooling law begins at age six, which makes it safe to assume that individuals begin their education at that age.

an individual faces no break in his or her education voyage. That is, he or she continues his or her education immediately upon arriving the United States and has not repeated a grade. In addition, assume that the effect of an additional year of U.S. education to be identical across individuals, regardless of what their ages at arrival are. That is, for an individual who arrived upon elementary graduation, the returns to studying an extra year in the United States are similar to another individual who arrived upon finishing third grade in his or her home country.

1.7.4 English Proficiency

I based the English proficiency measure on the self-reported Census question on *SPEAKENG*, which asks individuals to rank themselves as speaking English "not at all, some, well, very well." The indicator variable, *English*, receives a value of one for those who fall in the last three categories to the above question: some, well, or very well. For those who do not speak English at all, *English* is assigned a value of zero. Data obtained are imperfect measures of an immigrant's English skill for the following reasons. First, this Census question mainly asks about an individual's *current* English ability, instead of the language proficiency when he or she first arrived the United States. To truly capture the effect of English ability, it would be ideal to collect data on initial English fluency. Second, as for other self-reporting answers, measurement errors can occur. An individual may over- or under-state his or her English proficiency either intentionally or unintentionally. Since these cannot be corrected using the given data, the *English* ability variable is constructed as is given in the Census.

Chapter 2

The Use of Language and Culture: Does Speaking a Non-English Native Language Hurt or Benefit Immigrant Wages?

2.1 Introduction

Factors that influence wage rates have been one of the central interests in the immigrant research field. Many researchers have found that, due to their lack of English fluency, immigrants earn a lower wage rate compared to native workers (Borjas (1987, 1999, 2001); Chiswick and Miller (1995, 1998)). Although it is true that immigrants who are from a non-English-speaking country have a disadvantage in English proficiency, they have this asset of speaking a non-English language, such as fluency in a foreign language and knowledge about the goods and services market in other countries, which may enable them to be more productive at their jobs. If this hypothesis holds, then immigrants should be rewarded for this ability that distinguishes them from other native workers. In the literature, not much effort has been devoted to study this effect of a non-English language and cultural fluency on immigrant wages. To fill this gap, this paper explores aspects in which the detrimental effect of English deficiency on immigrant wages might be counteracted by positive labor market returns to speaking another language. In particular, through studying the value of language and knowledge in a foreign culture and industries where trade is a major activity in the business, a new perspective on factors influencing immigrant wages will be derived.

The first focus of this paper is to observe the effect of job requirements and the match with firms on immigrant wages. That is, immigrants who work at firms that value a foreign language and cultural skill should be paid more compared to those with similar attributes working in firms that do not have a need to use a foreign language. Specifically, does an immigrant who works in an area with heavy human-interaction on a daily basis earn higher wages compared to his counterpart who works in a field that values language to a lesser extent?

There are two possible channels through which proficiency in a non-English language can be beneficial to an immigrant. First, if an immigrant works in a job surrounded by individuals who share the same language and cultural background, the need to use English might be so low that they would choose to use their mother tongue to communicate with each other. For instance, for a Chinese immigrant who works as a salesperson in an area inhabited mainly by Chinese immigrants, the ability to speak Chinese readily distinguishes him from other native workers who are better at speaking English. Not only would this Chinese salesperson be able to communicate with a potential Chinese customer in their native language, but he may also understand the needs of an individual from the same cultural background, thus being better able to serve the customer. In addition, if an immigrant works with coworkers who share the same language background, they might choose to use their native language to communicate in the workplace. In this case, although English is the main language used in the United States, it might not also be the language of choice among this group of workers. If workers are able to initiate better communication at work, which enhances their productivity, then they may be compensated with a higher wage rate for this increase in productivity. This concept is similar to Borjas' (1991) notion of "ethnic capital," which is uniquely present among immigrants according to their ethnic culture. In this paper, the ability and choice of speaking one's mother tongue with other immigrants from the same cultural background can be viewed as a type of ethnic capital. With this ethnic capital, English might not necessarily be the sole language of choice used by immigrant workers.

An interesting note to the perspective suggested above concerns the effect of an increase in the pool of labor which possesses a specific skill in foreign language and culture. Holding demand constant, when there is a larger supply of labor who speaks the same non-English language, the equilibrium wage would be driven down. However, demand could also increase as a result, since the ability to speak this foreign tongue could further induce a greater customer base, which, in turn, increases the productivity of workers who speak this non-English language, therefore equilibrium wages paid could be driven up as well. As a result, there is no definite conclusion on how fluency in a foreign language and culture could affect an immigrant's wages earned. This issue could be addressed by the empirical results, which would be further explored in the next section.

Given the large number of immigrants living in the United States, whose language and cultural ability could be valued by firms that have connections with trading partners in a foreign country, the second purpose of this paper is to study whether working in an industry where international trade is large would contribute to the wages earned by immigrants. For example, the demand for a non-English language fluency could be high in the trade sector, particularly for firms with partners from foreign countries that may not adapt English as their official language(s). For example, Gould (1994) inspects the existence of an "immigrant information effect," which explains how fluency in a foreign language reduces trading costs when that language is indeed the official language of the trading partner's country. Using the log of U.S. exports of goods to one's home country as the outcome variable, he finds that the proportion of immigrants, relative to the total population in the United States, is predicted to positively affect bilateral trade volume. Therefore, hiring immigrant workers would reduce the firm's cost, since not only could the workers provide manpower, but they could also serve as an interpreter when the firm deals with its foreign trading partner. Additionally, immigrants who have lived in their home country for years would gain information on what kind of goods their country would be efficient in producing, which is a valuable piece of information to a firm. Furthermore, not only would immigrants have better information on the type of goods that their home country has a comparative advantage in producing, but they also possess information on the type of goods demanded, which would potentially enhance trading opportunities (Rauch and Trindade (1999); Chiswick and Miller (2002)). As a result, if immigrant workers could provide more information on potential trading opportunities to the firm, they would then receive a higher wage rate in return.

Although the above situations illustrate how the ability to speak a non-English language would be beneficial to the firm, the wage rate that a firm offers to an immigrant worker may be bounded if there is already an ample supply of that type of workers. In particular, holding labor demand constant, if there are a lot of immigrants who share the same language background in the area where the firm is located, the benefits of knowing a foreign language might cease, since there is a large supply of labor, holding other attributes constant. However, if the language ability of an immigrant is unique (such as having only a few German-speaking immigrants working in a region heavily occupied by Spanish-speaking immigrants), then the associated language skill is scarce, thus raises its value from the firm's perspective. As a result, the firm might be willing to offer a higher compensation in order to attract this worker. In other words, concentration of immigrants who share the same language background may also affect the returns to knowing a language.

Regression results show that, compared to another immigrant who has a job in another field, having a human-interaction-intensive job is associated with a positive wage gain of 4.47% on average. Regarding immigrant concentration, an increase in the percent of immigrants born in the same country as the individual under study is predicted to negatively affect his wages, while an increase in the percent of immigrants born in another country seems to enhance his wage rate. This addresses the issue that an increase in labor force that shares a common non-English language may drive the equilibrium wages down (holding labor demand constant) or up (when the ability to speak a foreign language increases the potential customer base that an immigrant worker can serve, which increases labor demand). Judging from the results, it seems that the former force carries a stronger influence compared to the latter. Furthermore, trade volume appears to positively affect the returns to working in the Wholesale Trade and the Transportation and Warehousing industries, on average. Specifically, for an immigrant worker serving in the Wholesale Trade sector, a one standard deviation increase in the bilateral trade volume between the United States and his home country is predicted to increase his wages by 3.36%[1].

This paper is organized as follows. Section 2.2 outlines a theoretical model that suggests an empirical framework, while section 2.3 provides details on the data sets used. The empirical strategy adopted is then described in section 2.4. Regression results are discussed in section 2.5. In section 2.6, additional checks are performed. Lastly, concluding remarks are given in section 2.7.

2.2 Theoretical Model and Hypotheses

To provide further understanding on the possibility that non-English language proficiency affects immigrant wages, a theoretical model based on the one developed in Bacolod, Blum, and Strange (2007) is described in this section.

As discussed in Lucas (1977), a worker's wage function can be treated as a hedonic equation, where wages are determined according to each of the characteristics that such worker possesses. Following this logic to study situations in which proficiency in a non-English language affects wages earned, it means that, for a firm that requires language as an input in its production function, the ability of a worker to speak a particular language will be taken into account in the wage determination process. In Bacolod, Blum, and Strange's (2007) model, which aims at explaining the effect of agglomeration in a city on worker i's productivity, marginal product of labor is dependent on the match between workers

[1] A separate set of regressions is run on individuals working in the manufacturing industry, with the reasoning that it is a sector that encounters verbal communication and foreign partnership more frequently. In manufacturing, workers may use their more fluent language to communicate in the workplace. In addition, import and export activities are more likely to take place, which induce a need for interpreters and someone who is more aware of foreign trading opportunities. Therefore firms would be willing to compensate qualified workers with a higher wage rate. However, regression results are not that different compared to the main results. In addition, the coefficient estimates on the *Trade volume* variable are consistently insignificant across all specifications. Therefore the associated estimates are not reported.

and firms located in various urban synergies. Applying this framework to discuss why two workers with identical skills set might receive different wages, the marginal productivity of labor for a worker with foreign-language and cultural ability is influenced by whether he works in a firm that requires such ability or not. That is, if this worker works in a firm that requires the usage of a non-English language (for instance, working in a trading firm), then he should receive a higher wage rate compared to a worker who works in another firm. In other words, the match between the firm and the worker is important.

Marginal product of labor, *MP*, is represented as:

$$MP = A + \delta^* a \tag{2.1}$$

where	A	=	worker's physical MP, excluding language ability
	δ	=	probability of a language match between the firm and the worker
	a	=	inherent value of a match with the firm

In other words, worker i's marginal product of labor depends not only on his working ability (A), but also on whether his language ability fits the need of the firm ($\delta^* a$), which depends on the probability of the job match between the firm and the worker. Equalizing marginal product of labor with wages, $w(z_i)$, which depends on the skill set, z_i, possessed by worker i, one gets

$$w(z_i) = A + \delta^* a \tag{2.2}$$

The above model suggests that the composite set of skills a worker possesses has an impact on his subsequent wages. In the market, if the worker (supplier) is endowed with a particular skill (such as foreign language fluency and cultural knowledge) that is of high value to the firm (demander), then the associated marginal product of labor would induce the firm to pay a higher wage rate to compensate for this additional skill unavailable in other workers. In addition, the match between the worker and the firm, defined by the set of skills required by the firm and that supplied by the worker, influences the worker's wages as well. Not only would a better match enhance worker productivity, which benefits the firm, but the worker would also receive a higher wage rate to compensate for possessing a skill not found in other candidates. Since mastery in a non-English language is usually one of the key elements that distinguishes an immigrant worker from a native worker, for a firm that requires this skill, an immigrant worker would subsequently earn a higher wage rate if he works in the firm, compared to another immigrant who has the same language ability but works in another firm that does not value foreign language fluency. An example of the type of firm that pays more for a worker with a language skill could be firms in the trade or sales industry, where human interactions take place at a heavier level. As a result, the model suggests that, for an immigrant worker whose mother tongue is non-English, working in an industry that values language usage would carry a positive effect on wages.

Under a competitive labor market, an excess supply of labor would induce a lower equilibrium wage rate. What this means is that, when there is a limited number of jobs that require language usage but with a constant inflow of labors who are proficient in that language, there should not be a difference in wages offered. As a result, even though a worker possesses a skill that suits the particular need of a firm, he would not receive a higher wage rate under a competitive labor market. However, studies in the literature (for example, Krueger and Summers (1988)) have argued for the existence of differential wages observed across industries, such as paying workers wages according to their distinctive characteristics to retain workers from leaving the industry. As a result, an immigrant worker who is fluent in speaking a language needed by a firm would still receive a higher rate of return compared to another worker who has similar attributes, but works at a job that does not require the use of language.

One of the main hypothesis tests performed will be that, for an immigrant, living in an area occupied by others with a similar language background and working in a profession where communication skill is highly valued, being proficient in his native language could carry an ambiguous effect on wages, compared to another immigrant with a similar background who works in another occupation. For example, if there is an increase in immigrants possessing similar language and cultural background, then, holding labor demand constant, having more immigrants entering the labor market would shifts labor supply to the right (as shown in Figure 2.1), which results in a decrease in wages. However, an increase in the number of immigrants who possess the same language and cultural background would also potentially increase labor demand, since these immigrants could be able to attract a wider customer base, especially if these customers came from the same country. As a result, holding labor supply constant, labor demand would increase (as shown in Figure 2.2), thus increases the equilibrium wages paid in the market. Since both labor supply and labor demand could shift, the result of having more immigrants entering the labor market on wages paid could be ambiguous (Figure 2.3).

Moreover, previous studies have found the tendency of clustering among immigrants (Lazear (1998; 1999)). Although living in an area filled with individuals who do not speak English well would hinder an immigrant's rate of English adaptation, it could, at the same time, create business opportunities for the immigrant. For instance, if this immigrant worker happens to hold a job that requires a heavy language usage, such as being a salesperson, the ability to speak a common non-English language would allow this immigrant to attract a broader customer base in the language enclave, for there is not a clear need to use English during the transaction. To test this hypothesis, a variable capturing percent of immigrants who came from the same country of origin will be incorporated in the regression framework. Furthermore, under the standard labor supply model with homogeneous workers, an increase in supply would reduce the equilibrium wage rate, holding demand constant. By testing whether the percent of immigrants residing in a particular county would positively or negatively affect the returns to working in an area with an intensive language usage, one may be able to infer whether immigrant workers who came from the same country are substitutes or complements to each other.

In order to draw valid conclusions, it would be ideal to control for ability and family background effects. However, these two factors remain unobserved in the data set, which means the coefficient estimates on the occupation-related variable are probably biased in all of the regression results.

2.3 Data

Two data sets will be used to perform the above hypothesis tests. To obtain information on individual characteristics, including country of birth, average hourly labor wages, English proficiency, gender, race, marital status, age, occupation, industry, residential location, and years in the U.S., the 5% Integrated Public Use Microsample Series (IPUMS) version of the 2000 Census will be used. Although the Census contains the majority of information needed, it has several disadvantages. First, it does not contain a direct measure on whether an individual speaks a non-English language. Second, there is also no direct measure on a non-English language fluency contained in the Census. These problems are resolved by matching the individual's place of birth with the main language(s) used in the respective country. Through using **The World Factbook 2007**, published by the Central Intelligence Agency, official language(s) of various countries can be obtained. Combining this information with the assumption that an individual would be fluent in the official language(s) used in his home country, the mother tongue of this individual can be determined. This specification assumes there is no variation in language fluency, but only variation in the returns to acquiring that language according to the match between language needs and the occupation and industry that the individual works in.

A measure on the percentage of immigrants living in the same area as the individual under study is based on the *city* variable provided in the Census, which is then used to derive the county in which the immigrant lives in. To precisely calculate a county-level immigrant concentration measure, statistics of individual states are obtained from the Census web page[2]. This web page provides information on the total population of each county in a U.S. state, as well as the number of foreign-born individuals (grouped at a country level) residing in that particular county. Through dividing the number of foreign-born by total population, a county-level immigrant concentration measure is obtained[3]. The advantage of measuring immigrant concentration at the county level is that an individual is more likely to work in the same county as his county, rather than city, of residence.

To test the effect of bilateral trade volume on wages earned by immigrants working

[2] *http://www.census.gov/qfd/states/*

[3] For details on the derivation, see the Appendix.

in the trade industry, the 1972 to 2001 import and export data created by Feenstra (1996) and Feenstra, Romalis, and Schott (2002)[4], which capture the volume of trade between the United States and various countries at an industry level, are used[5]. As a result of the need to match up the industry codes listed in the Census and the Feenstra (1996) and Feenstra, Romalis, and Schott (2002) data, the trade volume variable used is classified by industry. However, these two industry classifications differ substantially, therefore the sum of import and export volume is created as a single trade volume variable used in regressions presented in this paper, which is a national measure identical for everyone who came from the same country. Further, this measure is standardized to have a zero-mean with a standard deviation of one.

2.4 Empirical Model

In order to observe the effect of language and cultural proficiency on immigrant wages, the following empirical framework will be utilized:

$$y_{cijk} = \alpha_1 + X_i'\beta + \gamma_1 MP_{cik} + \alpha_k + \alpha_c + Z'\gamma + \epsilon_{cijk} \quad (2.3)$$

In the above model, y_{cijk} represents the log hourly wage for individual i who speaks home country's language c, works in occupation j, and lives in county k in the United States. X is a vector signifying individual characteristics, such as age, educational attainment, gender, years since migration (and its square term), marital status, and two dummy variables, one set to unity if the individual came as a child immigrant (i.e., before the age of 14), while the other is set to one if the person speaks English.

[4]Imports data are obtained from *http://cid.econ.ucdavis.edu/data/sasstata/usiss.html*, while exports information are found on *http://cid.econ.ucdavis.edu/data/sasstata/usxss.html*.

[5]Note that the *industry* variable in the Census and the industry information given in Feenstra's data sets are based on different classification systems. While the Census uses the *North American Industry Classification System* (NAICS) codes, Feenstra uses the *Standard Industrial Classification* (SIC) classification standard. This creates some difficulty in matching the data sets. In particular, after matching, there would only be three industries listed in the data set. They are agriculture, mining, and manufacturing. Details on how the codes are matched are discussed in the Appendix.

The main variables of interest are captured in the MP_{cik} term, which includes factors that affect worker i's marginal productivity. One example would be the match between worker characteristics and that required for the job. For instance, if mastery of a non-English language is valued at the job and this worker does indeed speak that language fluently, then his ability to speak this foreign language would enhance his marginal productivity. To study this effect, a dummy variable, $Human_{ij}$, will be set equal to one if individual i holds a job that requires more interpersonal communication skills, such as being a salesperson, a legal advisor, or a health practitioner[6]. That is, compared to another immigrant with similar observable attributes, would holding an occupation that involves more frequent human-interaction on a daily basis make a difference in wages earned by individual i?

Secondly, according to the standard labor model, holding demand constant, a shift in labor supply would also affect wages. To test whether immigrants who came from a different country of origin are viewed as complements or substitutes, *Percent Different Country* is incorporated in the model. More specifically, *Percent Different Country* is defined as the number of immigrants who originated from a country that is different from individual i's, divided by the total foreign-born population in a particular county that individual i resides in. By interacting this measure with $Human_{ij}$, the impact of an increase in labor supply on wages received by an immigrant who works in a human-interaction-intensive occupation can be studied. More specifically, through observing the coefficient estimate of this variable, one can deduce the substitutability of labor. That is, if the correlation between this variable representing the percent of other workers who share a different cultural and language background from the individual under observation and individual i's wages is negative, an increase in the number of workers from that particular group would reduce wages earned by i. Holding demand and other factors constant, workers belonging to that group could be substitutes to i.

Under the current framework, it is possible that other factors related to the characteristics of one's county of residence and occupational choice are omitted. For instance, a computer programmer might choose to live in a county in northern California, due to

[6]For more information on occupations classified as human-interaction-intensive, see the Appendix.

the number of related firms clustering in the Silicon Valley. As a result, failure to take county characteristics into account might result in a biased estimate on terms related to one's occupation. To capture the effects of unobserved county factors, a set of dummy variables representing an individual's county of residence, α_k, is created. As in many other immigrant wage rate studies, unobserved characteristics of one's country of origin could potentially be related to the individual's occupation or industry choice as well, thus biasing the coefficient estimate of industry-related regressors. As a remedy to this problem, a set of country of origin fixed effects dummies, α_c, is also incorporated in the regression model.

To test whether bilateral trade volume between the U.S. and one's country of origin affects the returns to working in the trade industry, Z represents additional variables of interest, such as the 1972 to 2001 import and export trade volume, based on Feenstra's (1996) and Feenstra, Romalis, and Schott's (2002) data sets. In this paper, trade volume, $Trade_c$, measured at an international level, is defined as the sum of imports and exports quantity that an individual country trades with the United States[7]. To test whether $Trade_c$ alters the returns to working in the trade industry, it will be interacted with one's industry of work. One of the main coefficients of interest would then be the interaction between $Trade_c$ and a dummy variable signifying the Wholesale Trade industry, for this is probably the industry with the most need of using knowledge in a foreign language and culture in order to facilitate business transactions.

2.5 Regression Results

2.5.1 The Relationship between Working in a Human-Interaction-Intensive Occupation and Immigrant Concentration

One of the major differences between an immigrant and a native worker is their respective language ability. As discussed previously, the effect of an immigrant's fluency in a non-English language on wages may be positive if language use is valued highly in the

[7] This measure is standardized to be mean-zero with a standard deviation of one.

immigrant's job need. To study whether this hypothesis holds true, a dummy variable, *Human*, is created to signify jobs that are more human-interaction intensive. For instance, *Human* is set to equal one for those who work as a salesperson, zero for those who are computer programmers. Regression results are recorded in Table 2.1.

As shown in Table 2.1, English ability and educational attainment are predicted to positively affect wages, which agrees with what has been found in other research work. An interesting trend to note is that coming to the United States at a younger age (before 14) seems to adversely affect wages in the sample used in this paper.

To study whether an immigrant's foreign language skill and cultural knowledge affect his wages earned when he works at a job that uses language more frequently, note that, holding country of origin and county of residence fixed, on average, an immigrant who has a job that requires a frequent need of communication is predicted to enjoy a wage gain of 4.47% (column (4)), compared to another immigrant whose job has less of an interaction need. In other words, holding other attributes, including English fluency, constant, working in an area that involves more usage of language is projected to affect immigrant wages positively. Since immigrants are proficient in a non-English language in general, would the percentage of immigrants who live in the same county further affect one's returns associated with working in a language-intensive job? To address this issue, an immigrant concentration measure, *Percent Different Country*, which signifies the percentage of total immigrant population who came from a different country of origin but also live in the same county as i, is created. Results obtained are listed in Table 2.2.

In Table 2.2, variable *Percent Different Country*[8] represents the total percentage of all immigrants living in the county but came from a different country of origin compared to individual i. Assuming i works in the same area as where he lives, an increase in the number of individuals who speak a different foreign language might reduce his ability to attract more customers, thus adversely affect his wages. On the other hand, if immigrants who share the same language are homogeneous, holding other factors, including labor demand, constant, an increase in the supply of labor with a different background would result in a

[8]All concentration measures included in this paper, such as *Percent Foreign Born* and *Percent Different Country*, are standardized to be mean-zero with a standard deviation of one.

raise in the equilibrium wage rate, for this language ability is uniquely found in i. According to regression results, the second scenario seems to dominate, since a one standard deviation increase in the percent of immigrants who came from a different country is predicted to increase an immigrant's wages by 1.75 % (column (3)). Also, as shown in column (4), the percent of immigrants born in a different country does not affect the returns to working in an interaction-intensive job for an immigrant. However, for those whose job has a greater language component, a one standard deviation increase in the total percentage of foreign-born in a particular county is predicted to increase wages by 4.71%. This may suggest that there is a psychological factor contributing to an immigrant's consumption behavior. That is, an immigrant may feel more comfortable obtaining goods and services from another immigrant instead of a native English speaker, regardless of this immigrant's own native language. If this is the case, then the wage rate earned by an immigrant who serves in a language-intensive job might be positively correlated to the number of immigrants living in the county.

Moreover, although the above findings suggest immigrants who possess a different language and cultural background are perceived as complements, those who work in an occupation with a higher probability of language usage are still at an advantage in terms of wages earned. Compared to another immigrant who has a job that requires less language usage, wages earned by an immigrant with a more human-interaction-intensive occupation would be 4.28% higher (column (3)).

Holding observable characteristics, such as age and educational attainment, constant, these findings can be attributed to the value of language and cultural skills perceived by the firm. For instance, if there are more Chinese workers available in a certain county, holding demand constant, a firm with a need for Chinese language usage in its production could easily find substitutes for its workers. In other words, the value of knowing Chinese is lower, so the firm does not need to pay a premium to workers who speak Chinese in the firm. On the other hand, if this firm has a demand for Spanish-speaking workers, with a high Chinese immigrant concentration, the ability to speak Spanish is more valuable, thus induces the firm to pay more to attract and retain Spanish-speaking workers. As a result, those who came from the same country are perceived as substitutes, while those who emigrated from a different country are viewed as complements.

2.5.2 The Effect of Trade

As suggested in the introduction, not only can an immigrant provide information on the goods market in his home country, which creates more trading opportunities for the firm, but he can also serve as an interpreter during the business process. In either case, bilateral trade volume between the United States and one's country of origin might also affect wages earned by the immigrant worker. To test this hypothesis, this section concentrates on whether trade volume alters the effect of working in an industry with a higher probability to trade with foreign partners on wages. Regression results are shown in Table 2.3[9].

If trade volume were to matter, it would most likely affect the returns to working in the trade industry, where it has a direct effect on the value that an immigrant worker brings to the firm. As shown from the coefficient of *Trade volume*Wholesale Trade* in column (2) of Table 2.3, for an immigrant working in the Wholesale Trade sector, a one standard deviation increase in bilateral trade volume between the U.S. and his home country is predicted to increase his wages by 3.36%. On a related note, for an immigrant serving in the Transportation and Warehousing sector, where business opportunities could be enhanced with an increase in foreign trading partners, a one-unit increase in trade volume would induce a 3.64% increase in wages.

Although trade volume seems to affect the returns to working in industries that have business connections with foreign countries, two interesting observations are obtained from column (2), with respect to an individual working in the Professional, Scientific, and Management, or the Arts, Entertainment, and Recreation field. When trade volume increases by one standard deviation, the wages for an immigrant worker associated with the Professional, Scientific, and Management industry are predicted to decrease by 7.28%. One possible explanation is the amount of outsourcing currently utilized by firms in the U.S. Since R&D is a common activity that takes place in this industry, outsourcing to

[9] In Table 2.3, those who work in the public sector are the omitted group. Also, trade volume and industry dummies are included in all of the regressions shown.

other foreign countries would reduce the value of foreign language ability among a local immigrant worker, for there is an abundant supply of workers who are proficient in both English and a non-English language in other parts of the world, where outsourcing takes place. Also, if major scientific research works are communicated in English and if fluency in a non-English language means less proficiency in English, the inability of an immigrant worker to use fluent English might be considered as a disadvantage in this field. As a result, an increase in trade volume between the United States and countries where firms based their outsourcing activity on could decrease wages earned by a local U.S. immigrant worker.

On the other hand, for a person working in the Arts, Entertainment, and Recreation industry, a one standard deviation increase in bilateral trade volume is associated with a 3.17% increase in wages earned by a local immigrant. This can be explained if the immigrant is viewed as a readily-available source to serve the industry's interest in foreign culture. That is, when a firm in the industry cooperates with foreign partners in producing entertainment products, a local immigrant worker who has already gained exposure to the respective country possesses a unique cultural knowledge that is valuable to the U.S. firm. To compensate for the value that this worker brings, the firm would subsequently offer a higher wage rate in return.

2.6 Additional Tests

2.6.1 Using the Subsample of Individuals who Came from a Spanish-Speaking Country

So far, an immigrant worker is seen as a liaison between the United States and his home country. However, from the language perspective, the value that an immigrant worker would add to the firm is not only bounded by his ability to communicate with others in his home country. That is, his ability to communicate with foreign trading partners who use the same mother tongue should also be taken into account. In this subsection, the sample is restricted to include only those who came from a Spanish-speaking country[10]. In this case, the potential value of a Mexican immigrant worker is evaluated by his ability to communicate with other overseas customers who live in Spanish-speaking countries[11]. Regression results are provided in Table 2.4.

Comparing column (4) between Tables 2.2 and 2.4, the returns to working in a language-usage-intensive job are positive and statistically significant. However, the magnitude of the coefficient estimate on *Human* becomes larger on the Spanish-speaking immigrant sample. Compared to another immigrant who does not have a job that requires more human interaction, the wages earned by someone who does could be 7.78% higher (contrast to 4.28% shown in Table 2.2). Another interesting observation is on the *Human*Percent Different Country*[12] term. For a Spanish-speaking immigrant who works at a job that involves more interaction with others, a one standard deviation increase in the percent of immigrants born from a different country is predicted to increase wages by 1.14%. This finding is different from that recorded when the whole sample is used. It is possible that, since Spanish is commonly used in many countries across the globe, the value that a Spanish-speaking worker brings to the firm is especially large. Therefore when there are less Spanish-speaking workers available in a county (that is, more non-Spanish-speaking immigrant workers), firms are willing to offer a higher wage rate to retain the worker with Spanish language ability who holds a job with heavy human-interactions. In addition, this finding suggests that the demand-side effect is greater than the supply-side effect. That is, although having more Spanish-speaking workers in the labor market could reduce the equilibrium wages paid (holding everything else constant), there is something unique among those who came from the same country of origin. For example, if there is a stronger cultural-tie among individuals from the same country, then an increase in the percent of

[10]It would be ideal to study the effect of language fluency across countries that use i's language, but since most of the languages used in a particular country is less universal, which greatly limit variations available, only Spanish-speaking countries are used to carry out this robustness test.

[11]These countries include Argentina, Bolivia, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Peru, Spain, and Venezuela.

[12]In here, *Percent Different Country* refers to the percentage of immigrants who came from a different Spanish-speaking country compared to individual i. For example, if individual i's country of origin is Spain, then *Percent Different Country* is defined as the percent of immigrants who came from a Spanish-speaking country that is not Spain and who reside in the same county as individual i.

immigrants born in another Spanish-speaking country would not affect an individual's desire to do business with another person who share the same cultural background. In this case, an increase in the percentage of immigrants born in a different country could make the individual's cultural value increase, and thus would affect his or her wages positively.

Also, as evident in Table 2.5, for a Spanish-speaking worker who works in the Professional, Scientific, and Management industry, a one standard deviation increase in bilateral trade volume is predicted to adversely affect wages. As discussed before, this finding can be explained by English being the universal language used in research work. As a result, if a Spanish-speaking worker is not as proficient in English compared to others, the increase in activities in this industry may negatively affect his wages.

2.6.2 Immigrants from English-Speaking Countries

Holding other attributes constant, the biggest difference between a native and an immigrant worker is English ability. However, an immigrant who came from a country that readily utilizes English as (one of) its official language(s) may be viewed as being similar to a native worker, for their fluency in the English language could be almost identical. As discussed in the introduction, the reasons why an immigrant worker adds value to the firm are because of his knowledge in the services demanded by individuals with a similar background, as well as his capability to serve as an interpreter when the firm trades with a foreign partner located in his country of origin. When one focuses on an immigrant originated from an English-speaking country[13], the effect of serving as a verbal interpreter on wage compensation may be lessened. However, an immigrant with an English mother tongue could still contribute to the firm by providing information on the cultural aspect of his country of origin (Borjas' (1991) ethnic capital concept). Also, if working in a human-interaction-intensive occupation still magnifies the effect of working in an immigrant enclave, then studying the wage pattern earned by this particular immigrant would build insight into how an immigrant worker may contribute to the firm's profit by understanding

[13]English-speaking countries include Australia, Barbados, Canada, Ghana, Hong Kong, Ireland, Jamaica, New Zealand, Nigeria, Pakistan, Philippines, Sierra Leone, Union of South Africa, Thailand, Trinidad and Tobago, and the United Kingdom, as defined in the CIA's **The World Factbook 2007**.

the needs of customers who came from the same country.

Concentrating on column (4) of Table 2.6, a one standard deviation increase in individuals born in a different country would induce a wage gain of 1.69% for an immigrant holding a job that encounters frequent human interactions, a finding that is very different from that recorded in Table 2.2. Applying the standard labor supply argument, the ability of an English-speaking worker to deliver cultural knowledge specific to his country of origin is especially valued when such knowledge is scarce in the area, especially when this worker works at a human-interaction-intensive job. Some other differences between results obtained when the whole sample and those who came from an English-speaking country are on the coefficient estimates on the *Percent Different Country* (column (3)) and *Human*Percent Foreign Born* (column (4)) terms. In Table 2.2, which uses the whole immigrant sample, for an immigrant worker with a job that uses language more frequently, a one standard deviation increase in the proportion of immigrants is associated with a 4.71% increase in wages. However, this effect diminishes if a worker immigrated from an English-speaking country (column (4), Table 2.6). Also, the percent of immigrants who originated from another English-speaking country is predicted to not affect individual i's wages, a finding that is different from previously found in column (3) of Table 2, when the whole sample was used. A possible explanation for these findings is that, judging from a language point of view, an immigrant who came from a country that uses English as an official language is not very different from a native worker. When there are more immigrants living in the area, this immigrant's ability to speak English would not induce other immigrants to do business with him, for they could have worked with a native worker should language not be a concern during the consumption process. If this the case, then an increase in the percentage of foreign born would not alter the returns to working at a human-interaction-intensive job for an immigrant who came from an English-speaking country.

With respect to whether trade volume alters the returns to working in a certain industry, figures are very different for an immigrant who came from an English-speaking country compared to someone from a country that does not use English as an official language. For instance, unlike those reported in Table 2.3, trade volume no longer affects the returns to working in the Wholesale Trade, Transportation and Warehousing, Professional, Scientific, and Management, nor Arts, Entertainment, and Recreation sectors. Instead, it is predicted to carry a positive effect on wages earned by an individual working in the Agriculture, Mining, Construction, or Manufacturing industry. It is possible that, since the majority of trading partners in these industries are English-speaking[14], the importance of being fluent in a non-English language is reduced. As a result, trade volume does not alter the returns to working in these industries for immigrants included in the original sample (Table 2.3), but its association with the wages earned by an English-speaking immigrant worker would be positive[15].

2.7 Conclusion

The main findings in this paper are three-fold. First, the correlation between having a job that requires heavy human-interaction on a regular basis and wages is positive, regardless of whether an immigrant came from a country that utilizes a non-English official language or not. Second, when the number of individuals with the same language background increases, the wage rate earned by an immigrant working with a human-interaction-intensive job would tend to decrease; when the number of workers who came from a different country increases, the associated wages earned by a particular immigrant would increase. This phenomenon implies that an immigrant's advantage in attracting more customers who share the same language background is not strong enough to induce a positive relationship between wages and the number of individuals born in the same country. That is, the demand-side effect is weaker than the supply-side argument, as shown in Figure 2.1. Thus, this finding can be explained using a standard labor supply model. When the skill possessed by a worker is more unique, firms who need that language in its production func-

[14] For example, Canada, the largest trading partner with the United States, is also the largest electricity importer from the U.S. in 2000, according to the Energy Information Administration (2005).

[15] As in many other studies, the presence of unobservables might cause bias in the coefficient estimates obtained. Depending on the correlation between such unobservables and the *Human* dummy variable, the estimated coefficient might be biased upward or downward. To explore the possibility of having one's country of origin being correlated with his county of residence, a set of regression that incorporates country of origin × county of residence is performed. Since most of the results are similar to that reported in Table 2.2, they are not reported here.

tion would be willing to offer a higher premium to attract the worker. On the other hand, when such worker's skill becomes common, the induced increase in labor supply, holding demand constant, would drive the equilibrium wage rate downward. Third, trade volume is empirically predicted to magnify the returns associated with working in the Wholesale Trade industry for an immigrant. Not only would this immigrant understand the needs of the goods and services market in his home country, but during the trading process, he can also serve as an interpreter for the firm, which makes him more valuable compared to a native worker who is similar in other aspects.

In addition to the above findings, it is also interesting to note that the effect of trade volume on wages earned by an immigrant working in various industries differs, depending on whether the immigrant came from an English-speaking home country or not. In particular, trade volume is predicted to have no effect on an immigrant with an English language background and who works in the Wholesale Trade industry. Instead, it is positively associated with wages earned when the English-speaking individual works in industries such as Mining, where U.S. firms tend to trade with countries that use English as an official language. This, however, does not imply that an immigrant coming from an English-speaking country would contribute to the firm in the same manner as a native worker, because a foreign-born individual still possesses a unique cultural knowledge that cannot be found in a native worker.

To study the effect of bilateral trade volume on immigrant wages, it would be ideal to measure trade volume at a finer measure, such as by industry instead of by country. Due to the difference in industry classification between the Census and the Feenstra (1996) and Feenstra, Romalis, and Schott (2002) data, trade volume by industry level was not easy to obtain. It would be interesting to perform this study again with a refined trade volume measure, which can be achieved when data becomes available.

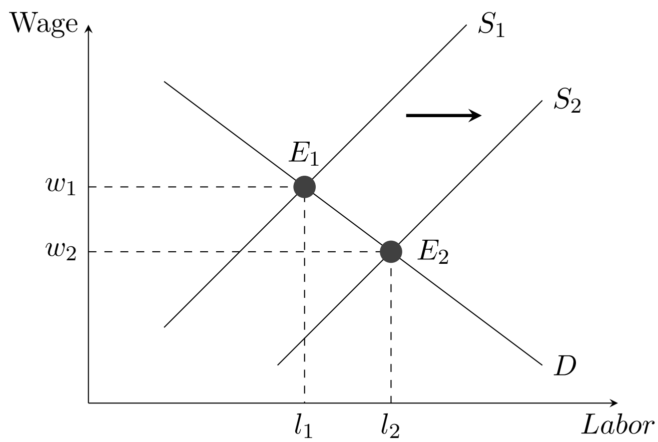

Figure 2.1: The effect of an increase in labor supply on equilibrium wages

When there is an increase in the number of immigrants who share the same language and cultural background, labor supply would shift out (from S_1 *to* S_2*, meaning that the number of workers available for a certain job that requires a specific language and cultural skill would also increase (shown by* l_2*, compared to* l_1*). Holding demand constant, the equilibrium wage rate would go down from* w_1 *to* w_2 *as well.*

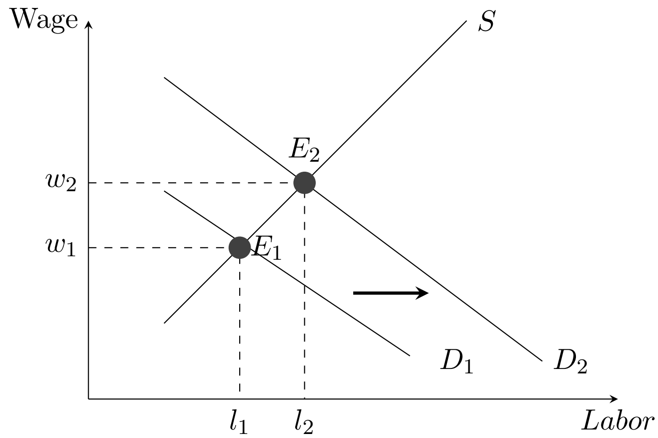

Figure 2.2: The effect of an increase in labor demand on equilibrium wages

However, when there are more workers who share the share language and cultural background available in the local labor market, their ability to communicate with other customers who share the same background could enable them to attract a larger customer base. In this case, labor demand would increase (shifting from D_1 to D_2). Holding labor supply constant, the equilibrium wage rate increase from w_1 to w_2, while the number of workers also increases from l_1 to l_2.

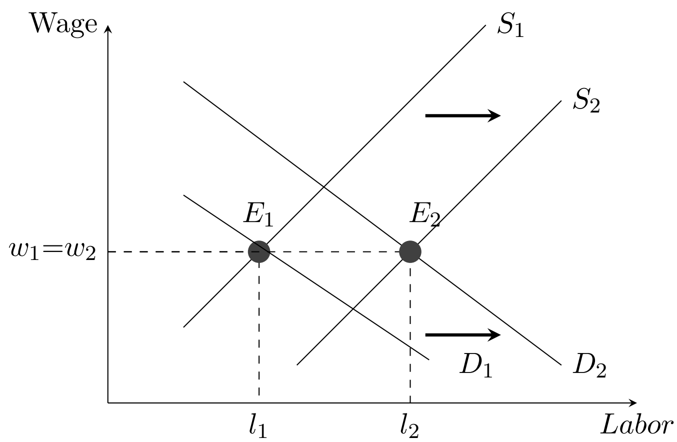

Figure 2.3: The effect of a simultaneous increase in labor demand and labor supply on equilibrium wages

When the effects demonstrated in Figures 2.1 and 2.2 happen simultaneously, their effects on the equilibrium wage rate could become ambiguous. For example, as shown in the above graph, wages could remain unchanged (w_1=w_2). If the magnitude of labor supply shift is greater than that of labor demand (not shown), wages could decrease, and vice versa. As a result, having more immigrant workers who speak the same language and share an identical cultural background might increase, decrease, or carry no effect on the equilibrium wage offered in the labor market.

Table 2.1: The Relationship between Working in an Occupation that Involves More Frequent Human Interaction and Immigrant Wages

	(1)	(2)	(3)	(4)
English Ability	0.1976*	0.1918*	0.1632*	0.1635*
	(0.0071)	(0.0071)	(0.0071)	(0.0071)
High School Graduate	0.1833*	0.1702*	0.1409*	0.1406*
	(0.0050)	(0.0050)	(0.0051)	(0.0051)
College Dropout	0.3763*	0.3660*	0.3093*	0.3118*
	(0.0055)	(0.0056)	(0.0059)	(0.0059)
College Graduate	0.7393*	0.7247*	0.6476*	0.6485*
	(0.0057)	(0.0059)	(0.0067)	(0.0067)
Child	-0.0834*	-0.0855*	-0.0953*	-0.0938*
	(0.0063)	(0.0063)	(0.0063)	(0.0063)
Human	0.0610*	0.0589*	0.0431*	0.0447*
	(0.0043)	(0.0043)	(0.0044)	(0.0045)
F(All county dummies jointly zero)		16.02		12.12
p-value		[0.0000]		[0.0000]
F(All country dummies jointly zero)			55.02	46.65
p-value			[0.0000]	[0.0000]
County Fixed Effects	No	Yes	No	Yes
Country Fixed Effects	No	No	Yes	Yes
N	139911	139911	139911	139911

[1] Regression results based on foreign-born immigrant workers aged 25 to 55 who came from a country with recorded bilateral trade volume between the United States and the respective country and who reside in a U.S. county where immigrant concentration data are available. Other regressors include dummy variables signifying educational attainment (whether the individual is a high school graduate, a college dropout, or a college graduate), a dummy variable signifying whether the person speaks English, another dummy variable set to unity for an individual who came as a child immigrant (before age 14), age, gender, years since migration (and its squared term), race, and marital status. County of residence and country of origin fixed effects dummies are included when applicable. Data source: 5% IPUMS version of Census 2000.

[2] * signifies 5% statistical significance. Robust standard errors shown in parentheses.

[3] *Human* is a dummy variable assigned a value of one for those who hold a job that involves an intensive human interaction on a daily basis. An example includes salesmanship. See the Appendix for more details.

Table 2.2: The Relationship between Percentage of Immigrants from a Different Country of Origin, Occupation, and Immigrant Wages

	(1)	(2)	(3)	(4)
Human	0.0447*	0.0428*	0.0445*	0.0428*
	(0.0044)	(0.0044)	(0.0044)	(0.0044)
Human*Percent Foreign Born		0.0471*		0.0471*
		(0.0039)		(0.0039)
Percent Different Country			0.0175*	
			(0.0052)	
Human*Percent Different Country				0.0002
				(0.0041)
F(All county dummies jointly zero)	12.12	11.03	11.56	10.96
p-value	[0.0000]	[0.0000]	[0.0000]	[0.0000]
F(All country dummies jointly zero)	46.65	46.44	43.96	45.39
p-value	[0.0000]	[0.0000]	[0.0000]	[0.0000]
County Fixed Effects	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes
N	139911	139911	139911	139911

[1] * signifies 5% statistical significance. Robust standard errors shown in parentheses.

[2] *Percent Different Country* represents the percentage of immigrants who came from a different country of origin but live in the same county as the individual observation. This variable has been standardized to have a zero mean with a standard deviation of one. Without normalization, the sum of *Percent Same Country* (percentage of immigrants who share the same country of origin) and *Percent Different Country* should be one. Due to their complementarity in the sign of coefficient, only regression results using *Percent Different Country* are reported.

Table 2.3: The Effect of Bilateral Trade Volume between the United States and an Immigrant's Country of Origin on Immigrant Wages

	(1)	(2)
Trade volume*Agriculture	-0.0532	-0.0426
	(0.0297)	(0.0319)
Trade volume*Mining	-0.0825	-0.0201
	(0.0599)	(0.0600)
Trade volume*Utilities	-0.0596*	-0.0097
	(0.0293)	(0.0317)
Trade volume*Construction	-0.0412*	0.0042
	(0.0064)	(0.0125)
Trade volume*Manufacturing	-0.0370*	-0.0092
	(0.0038)	(0.0115)
Trade volume*Wholesale Trade	-0.0035	0.0336*
	(0.0049)	(0.0120)
Trade volume*Transportation and Warehousing	-0.0052	0.0364*
	(0.0095)	(0.0146)
Trade volume*Information and Communications	-0.0495*	-0.0063
	(0.0132)	(0.0139)
Trade volume*Finance, Insurance, and Real Estate	-0.0640*	-0.0079
	(0.0089)	(0.0138)
Trade volume*Professional, Scientific, and Management	-0.1197*	-0.0728*
	(0.0078)	(0.0146)
Trade volume*Educational, Health, and Social Services	-0.0736*	-0.0181
	(0.0059)	(0.0123)
Trade volume*Arts, Entertainment, and Recreation	0.0030	0.0317*
	(0.0050)	(0.0120)
Trade volume*Other Services	-0.0236*	0.0120
	(0.0076)	(0.0133)
F(All county dummies jointly zero)		11.42
p-value		[0.0000]
F(All country dummies jointly zero)		37.41
p-value		[0.0000]
County Fixed Effects	No	Yes
Country Fixed Effects	No	Yes
N	139911	139911

[1] * signifies 5% statistical significance. Robust standard errors shown in parentheses.

Table 2.4: The Relationship between Percentage of Immigrants from a Different Country of Origin, Occupation, and Wages Earned by Immigrants From Spanish-Speaking Countries

	(1)	(2)	(3)	(4)
Human	0.0789*	0.0771*	0.0787*	0.0778*
	(0.0060)	(0.0061)	(0.0060)	(0.0060)
Human*Percent Foreign Born		0.0145*		0.0141*
		(0.0058)		(0.0058)
Percent Different Country			0.0074*	
			(0.0034)	
Human*Percent Different Country				0.0114*
				(0.0045)
F(All county dummies jointly zero)	7.53	7.61	7.46	7.62
p-value	[0.0000]	[0.0000]	[0.0000]	[0.0000]
F(All country dummies jointly zero)	12.35	12.38	12.27	12.87
p-value	[0.0000]	[0.0000]	[0.0000]	[0.0000]
County Fixed Effects	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes
N	72700	72700	72700	72700

[1] Regression results based on foreign-born immigrant workers aged 25 to 55 who came from a country that uses Spanish as its official language (Argentina, Bolivia, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Peru, Spain, and Venezuela) with relevant immigrant concentration data in the U.S. country of residence available. Other regressors include dummy variables signifying educational attainment (whether the individual is a high school graduate, a college dropout, or a college graduate), a dummy variable signifying whether the person speaks English, age, gender, years since migration (and its squared term), race, a dummy variable set to unity for those who came before age 14, and marital status. County of residence and country of origin fixed effects dummies are included when applicable. Data sources: 5% IPUMS version of Census 2000.

[2] * signifies 5% statistical significance. Robust standard errors shown in parentheses.

[3] *Percent Different Country* represents the percentage of immigrants who came from a different Spanish-speaking country of origin (Argentina, Bolivia, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Peru, Spain, or Venezuela) but live in the same county as the individual observation. This variable has been standardized to have a zero mean with a standard deviation of one. Data source: *http://www.census.gov/qfd/states/*.

[4] *Human* is a dummy variable assigned a value of one for those who hold a job that involves an intensive human interaction on a daily basis. An example includes salesmanship. See the Appendix for more details.

Table 2.5: The Effect of Bilateral Trade between the United States and Spanish-speaking Countries on Immigrant Wages

	(1)	(2)
Trade volume*Agriculture	-0.0205	-0.0130
	(0.0367)	(0.0385)
Trade volume*Mining	0.0939	0.0978
	(0.0696)	(0.0709)
Trade volume*Utilities	-0.0716*	-0.0317
	(0.0330)	(0.0363)
Trade volume*Construction	0.0043	0.0148
	(0.0083)	(0.0131)
Trade volume*Manufacturing	-0.0056	0.0034
	(0.0049)	(0.0112)
Trade volume*Wholesale Trade	-0.0042	0.0103
	(0.0061)	(0.0120)
Trade volume*Transportation and Warehousing	0.0139	0.0243
	(0.0112)	(0.0155)
Trade volume*Information and Communications	-0.0132	0.0079
	(0.0184)	(0.0211)
Trade volume*Finance, Insurance, and Real Estate	-0.0409*	-0.0200
	(0.0109)	(0.0151)
Trade volume*Professional, Scientific, and Management	-0.0677*	-0.0522*
	(0.0075)	(0.0127)
Trade volume*Educational, Health, and Social Services	-0.0456*	-0.0229
	(0.0071)	(0.0125)
Trade volume*Arts, Entertainment, and Recreation	-0.0206*	-0.0087
	(0.0062)	(0.0119)
Trade volume*Other Services	-0.0123	-0.0013
	(0.0088)	(0.0135)
F(All county dummies jointly zero)		7.32
p-value		[0.0000]
F(All country dummies jointly zero)		8.74
p-value		[0.0000]
County Fixed Effects	No	Yes
Country Fixed Effects	No	Yes
N	72700	72700

[1] * signifies 5% statistical significance. Robust standard errors shown in parentheses.

Table 2.6: The Relationship between Percentage of Immigrants from a Different Country of Origin, Occupation, and Wages Earned by Immigrants from English-Speaking Countries

	(1)	(2)	(3)	(4)
Human	0.0398*	0.0396*	0.0396*	0.0394*
	(0.0087)	(0.0087)	(0.0087)	(0.0087)
Human*Percent Foreign Born		0.0139		0.0098
		(0.0081)		(0.0082)
Percent Different Country			0.0070	
			(0.0059)	
Human*Percent Different Country				0.0169*
				(0.0067)
F(All county dummies jointly zero)	5.53	4.42	5.45	4.41
p-value	[0.0000]	[0.0000]	[0.0000]	[0.0000]
F(All country dummies jointly zero)	24.00	24.03	22.60	22.38
p-value	[0.0000]	[0.0000]	[0.0000]	[0.0000]
County Fixed Effects	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes
N	25841	25841	25841	25841

[1] Regression results based on foreign-born immigrant workers aged 25 to 55 who came from a country that uses English as (one of) its official language(s) with recorded bilateral trade volume between the United States and the respective country and who reside in a U.S. county where immigrant concentration data are available. Other regressors include dummy variables signifying educational attainment (whether the individual is a high school graduate, a college dropout, or a college graduate), age, gender, years since migration (and its squared term), race, a dummy variable set to unity for those who came before age 14, and marital status. County of residence and country of origin fixed effects dummies are included when applicable. Data source: 5% IPUMS version of Census 2000. Percentage of immigrants from respective countries who live in various counties in the U.S. is calculated using information obtained from *http://www.census.gov/qfd/states/*.

[2] * signifies 5% statistical significance. Robust standard errors shown in parentheses.

Table 2.7: The Effect of Bilateral Trade between the United States and the Immigrant's Country of Origin, which Adopts English as an Official Language

	(1)	(2)
Trade volume*Agriculture	0.2804	0.3493*
	(0.1449)	(0.1752)
Trade volume*Mining	0.1952*	0.1669*
	(0.0720)	(0.0838)
Trade volume*Utilities	0.0293	0.0314
	(0.0223)	(0.0268)
Trade volume*Construction	0.0302*	0.0370*
	(0.0148)	(0.0168)
Trade volume*Manufacturing	0.0542*	0.0487*
	(0.0135)	(0.0144)
Trade volume*Wholesale Trade	0.0227	0.0034
	(0.0211)	(0.0218)
Trade volume*Transportation and Warehousing	0.0119	0.0061
	(0.0369)	(0.0355)
Trade volume*Information and Communications	0.0002	-0.0107
	(0.0293)	(0.0303)
Trade volume*Finance, Insurance, and Real Estate	0.0251	0.0166
	(0.0190)	(0.0194)
Trade volume*Professional, Scientific, and Management	0.0196	0.0149
	(0.0116)	(0.0095)
Trade volume*Educational, Health, and Social Services	-0.0081	-0.0093
	(0.0055)	(0.0063)
Trade volume*Arts, Entertainment, and Recreation	0.0462	0.0386
	(0.0294)	(0.0300)
Trade volume*Other Services	-0.0003	0.0006
	(0.0039)	(0.0036)
F(All county dummies jointly zero)		5.36
p-value		[0.0000]
F(All country dummies jointly zero)		16.09
p-value		[0.0000]
County Fixed Effects	No	Yes
Country Fixed Effects	No	Yes
N	25841	25841

[1] * signifies 5% statistical significance. Robust standard errors shown in parentheses.

Table 2.8: Summary Statistics

Variable	Observations	Mean	Standard Deviation
ln (wage)	139911	2.2599	0.7834
English	139911	0.9147	0.2794
Male	139911	0.5863	0.4925
Married	139911	0.5781	0.4939
Years since Migration	139911	14.9138	9.8669
High School Graduate	139911	0.2554	0.4361
College Dropout	139911	0.2064	0.4048
College Graduate	139911	0.2416	0.4280
White	139911	0.3247	0.4683
Age	139911	35.9530	9.8127
Child	139911	0.2348	0.4239
Human	139911	0.3905	0.4879
Percent Same Country	139911	-0.0060	0.9990
Percent Different Country	139911	0.0060	0.9990
Trade volume	139911	-1.50×10^{-8}	1
Percent Foreign Born	139911	-4.37×10^{-9}	1

[1] *Child* is a dummy variable that refers to immigrants who came before the age of 14.

[2] *Human* is a dummy variable set to unity for occupations that involve a heavier human-language-interaction. Details on the variable construction can be found in the Appendix.

[3] *Percent Same Country*, *Percent Different Country*, *Trade volume*, and *Percent Foreign Born* are all standardized to be mean zero with a standard deviation of one. *Percent Same Country* and *Percent Different Country* represent the proportion of individuals who came from the same or different country of origin compared to immigrant *i*, respectively. *Trade volume* refers to the sum of imports and exports between the United States and various country between 1992 to 2001, as provided in Feenstra's trade data set based on Feenstra (1996) and Feenstra, Romalis, and Schott (2002). Lastly, *Percent Foreign Born* signifies the percent of immigrant living in a certain county. Data sources: 5% IPUMS version of Census 2000; Feenstra (1996) and Feenstra, Romalis, and Schott (2002); Percentage of immigrants from respective countries who live in various counties in the U.S. is calculated using information obtained from *http://www.census.gov/qfd/states/*.

2.8 Appendix

2.8.1 Countries Included

The sample includes immigrants who came from 64 different countries, including Afghanistan, Argentina, Australia, Austria, Barbados, Bangladesh, Bolivia, Cambodia, Canada, Chile, China, Colombia, Costa Rica, Cuba, Czech Republic, Dominican Republic, Ecuador, Egypt, El Salvador, France, Germany, Ghana, Guatemala, Haiti, Honduras, Hong Kong, Hungary, India, Indonesia, Iran, Iraq, Ireland, Israel, Italy, Jamaica, Japan, Korea, Laos, Lebanon, Mexico, Malaysia, Netherlands, New Zealand, Nicaragua, Nigeria, Pakistan, Panama, Peru, Philippines, Poland, Portugal, Romania, Sierra Leone, Spain, Sweden, Syria, Union of South Africa, Taiwan, Thailand, Trinidad and Tobago, Turkey, the United Kingdom, Venezuela, and Yugoslavia. Spanish-speaking countries include Argentina, Bolivia, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Peru, Spain, and Venezuela. Countries that use English as (one of) its official language(s) are Australia, Barbados, Canada, Ghana, Hong Kong, Ireland, Jamaica, New Zealand, Nigeria, Pakistan, Philippines, Sierra Leone, Union of South Africa, Thailand, Trinidad and Tobago, and the United Kingdom.

2.8.2 Immigrant Concentration

Instead of using an immigrant concentration variable measured at a state-level, this paper uses information provided on the U.S. Census web page[16] to calculate a county-level immigrant concentration measure. Specifically, the Census web page provides information on the number of individuals born in a certain country, as well as the total population in a particular county of a state. In addition, the web page contains information on county definition, which lists the name(s) of the city (cities) that belong(s) to that particular county. Through matching the individual observation's city of residence to its corresponding county (counties) and the total population in the corresponding county (counties), the

[16] *http://www.census.gov/qfd/states/.*

percent of individuals born from a particular non-U.S. country who reside in a county can be obtained. Note that, when a city is shown to locate at the boundary of multiple counties (such as the case for New York city, which is listed under the Bronx, Kings, New York, Queens, and Richmond counties), an average of the counties' total population is taken to derive the percentage of foreign-born living in that county. For instance, in the case of New York city, the percentage of immigrants born in Canada is calculated using the following formula:

$$\text{\% of immigrants born in Canada} = \frac{\text{Number of individuals born in Canada}}{0.20 \times A}$$

where A equals the sum of total population in Bronx, Kings, New York, Queens, and Richmond counties.

2.8.3 Industry

The industry dummy variables are defined according to the Census classification system, which is a set of 3-digit codes based on the 3- or 4-digit *North American Industry Classification System* (NAICS). Broader categories are also given in the Census, which combines similar elements in the industry classification into one general category. For example, under the more general *Agriculture, Forestry, Fishing and Hunting* group, there are six related industries listed (*Crop production*, *Animal production*, *Forestry except logging*, *Logging*, *Fishing, hunting, and trapping*, and *Support activities for agriculture and forestry*), each with a distinctive code (017, 018, 019, 027, 028, and 029, respectively). To match this information with the Feenstra (1996) and Feenstra, Romalis, and Schott (2002) data, which adopts a 1987 4-digit *Standard Industrial Classification* (SIC) system, the *1997 NAICS United States Structure, Including Relationship to 1987 U.S. SIC*, a publication released by the Census Bureau, is utilized. Although some of the NAICS codes provided in the reference contain 6 digits, meaning that industries are classified at a finer level, to make the reference useful in applying to the Census data set, only the first 4 digits are considered. Upon matching the NAICS codes with the SIC classification, trade volume in most of the industries can be identified. However, when there are ambiguities in the Feenstra or Feenstra el al. data (To gain further insight into the reasons for these ambiguities, see Feenstra (1996).), or when the NAICS codes cannot be easily matched to the SIC system, the related trade volumes are dropped from the sample.

Due to this complication in matching up industry definitions across three data sets, in all of the regressions ran, trade volume is grouped into three broad categories at an industry level: manufacturing, agriculture, and mining.

2.8.4 Occupation

General occupation classifications are taken from the Census definitions. In the 2000 Census, each occupation is assigned a specific occupation code. Similar to the industry codes, occupation codes are grouped under more general categories. For example, 13 of the distinct occupation codes are combined to form the *Financial Specialist* group. In this paper, only the broader occupation classification is used to create occupation dummy variables. However, for the *Human* dummy variable, which is assigned a value of 1 for occupations that involve a heavier human-interaction on a regular basis, a combination of the broad and detailed occupation classifications are used. Occupations defined as interaction-intensive include *Management, Business and Financial Operations*, *Business Operations Specialists*, *Financial Specialists*, *Computer Support Specialists*, *Architects, Except Naval*, *Surveyors, Cartographers, and Photogrammetrists*, *Community and Social Services*, *Legal*, *Education, Training, and Library*, *Actors*, *Producers and Directors*, *Athletes, Coaches, Umpires, and Related Workers*, *Announcers*, *News Analysts, Reporters, and Correspondents*, *Public Relations Specialists*, *Editors*, *Healthcare Practitioners and Technical*, *Healthcare Support*, *Protective Service*, *Bartenders*, *Combined Food Preparation and Serving Workers, Including Fast Food*, *Counter Attendants, Cafeteria, Food Concession, and Coffee Shop*, *Waiters and Waitresses*, *Food Servers, Nonrestaurant*, *Hosts and Hostesses, Restaurant, Lounge, and Coffee Shop*, *Tour and Travel Guides*, *Child Care Workers*, *Residential Advisors*, *Sales*, *Telephone Operators*, *Tellers*, *Customer Service Representatives*, *Hotel, Mo-*

tel, and Resort Desk Clerks, *Interviewers, Except Eligibility and Loan*, *Loan Interviewers and Clerks*, *Receptionists and Information Clerks*, *Reservation and Transportation Ticket Agents and Travel Clerks*, and *Secretaries and Administrative Assistants*.

2.8.5 A Country's Official Language

The World Factbook 2007, published by the CIA, contains detailed information on language(s) used in various countries. In most cases, it also lists explicitly which language is used as an official language in a particular country. Combining what is given in **The World Factbook 2007** and an individual's place of birth, variables signifying an individual's mother tongue can be identified. In case a country has more than one official languages, individuals originating from that country are assumed to be fluent in all of the official languages. For instance, since Chinese and English are both listed as the official languages in Hong Kong, immigrants coming from Hong Kong are, therefore, assumed to master both languages. As a result, the value for the language dummies *Chinese* and *English* are both set to be 1 for such immigrant.

Chapter 3

The Correlation between Past and Future: Does Parent Involvement in the Past Influence Student's Current Reading Performance?

3.1 Introduction

The purpose of this paper is to investigate whether exposure to activities related to reading at a younger age affects the performance in a reading test for fourth grade students. In general, factors that influence student performance in various tests have been one of the main areas of interest in the research field. Using measures such as class size and teacher credentials, analysis on the correlation between school inputs and student performance has been conducted. However, some research work, such as the Coleman Report (1960), have suggested that there is a stronger link between family background, compared to school resources, and test score. Perhaps due to data limitation, these studies have not been able to further decompose the effect of parental inputs on the child's outcome to a further extent. That is, which part of family inputs would affect the child's education outcome the most? With the use of the Progress in International Reading Literacy Study (PIRLS) data set, this paper aims at filling the missing piece of whether having a parent who would spend time engaging in activities related to learning how to read in the preschool years would carry any effect on a fourth grade child's reading performance.

Progress in International Reading Literacy Study (PIRLS) is a reading test given to fourth graders across at least 35 countries, in 2001 and 2006. So far, its use in the literature has been combined with two other international tests, Trends in International Mathematics and Science Study (TIMSS) and the Program of International Student Assessment (PISA), to perform cross-country analysis. For example, Schnepf (2007) studies whether language skill, socioeconomic status, and school segregation affect the differential performance between immigrants and natives in 10 OECD countries (Australia, Canada, France, Germany, the Netherlands, New Zealand, Sweden, Switzerland, the United Kingdom, and the United States). In addition, PIRLS has been used to study how peer effect affects a student's test score in various countries (Ammermüeller and Pischke (2006)). Other scholars, such as Hanushek and Woessmann (2005), combine the three international assessments to observe the effect of tracking on student performance, as well as on whether tracking induces some students to gain by making others lose. So far, other than Woessmann and Fuchs (2005), who study the relationship between family background[1] and student's PIRLS performance in Argentina and Colombia, not much has been done to directly address whether various aspects of parental inputs in a child's early life influence how well the student fares in the reading test, which is what this paper contributes. Specifically, this paper aims at using answers given in the parent survey to investigate if there is any correlation between time allocated to doing reading-related activities with a child and the child's test score, as recorded in the 2001 PIRLS assessment.

[1]While Woessmann and Fuchs (2005) use the number of books, as well as the parents' total educational attainment and occupation as family background variables, this paper uses whether the parent has participated in reading-related activities during the child's preschool years, annual family income, the amount of time the parent spends reading per week, and both of the parents' immigrant status as background measures. The only common variable used in Woessmann and Fuchs' (2005) and my paper is the parental educational attainment measure.

Administered by the International Association for the Evaluation of Educational Achievement (IEA), PIRLS is particularly useful in performing this analysis because, in addition to recording a nine year-old student's test score, it has also conducted surveys on the student, parent, teacher, and principal of the school. In particular, the parent survey contains detailed information on activities that the parent[2] has spent time participating with the student when he or she was younger (i.e., before formally started schooling). For instance, frequencies of parent(s) engaged in reading, telling stories, playing alphabet toys[3], writing words, or reading signs aloud with the student were recorded. Such questions provide a more detailed background on how the student was brought up, and thus supply useful information on studying the correlation between family background and student performance from a different angle.

In addition, this paper investigates whether there is a correlation between teacher characteristics and student performance in PIRLS, holding a student's prior reading ability constant. As suggested in Betts, Zau, and Rice (2003) and Hanushek (2002), using a framework that controls for prior performance to address how a specific education policy affects a student's outcome would be a reliable method to evaluate the effect of such policy on student performance. In the principal survey of PIRLS, principals are asked to estimate knowledge related to reading for their students before students started their first grade education. In particular, they report the average number of students who knew their alphabet or possessed the ability to write some words before entering first grade. Through the use of this part of the survey, the student's initial reading ability can be proxied and included in the regression, which would take the form of weighted least squares (WLS) and quantile regression, to study whether parental inputs during the child's early years would affect the child's performance in PIRLS.

To discuss whether parental inputs would have an effect on the child's reading test

[2]That is, the adult who filled out the survey. This could have been the student's (step-) father, (step-) mother, or a legal guardian.

[3]The existence of a relationship between playing board games and learning can be found in Cavanagh (2008), who reports findings from an experiment conducted by Siegler and Ramani at Carnegie Mellon and the University of Maryland, respectively, on preschool children. In particular, the use of board game in the classroom is found to positively influence a student's mathematics ability, especially if he or she comes from a family with a disadvantaged background.

score, it is important to first address the possibility of omitted variables bias. In particular, the endogeneity related to a parent's choice of the amount of time spent participating in various reading-related activities with the child. One of the factors that can influence how devoted a parent is in reading with the child is his or her socioeconomic status. For a parent with a higher SES, reading with the child to build up his or her interest in knowledge at an early age might be one of the main concerns. Also endogenous is the choice of school. Since a parent from a higher SES would possibly be exposed to more choices of school for his or her child by, for example, being able to afford the cost of moving to another jurisdiction where the schools would have a better quality compared to somewhere else[4]. If this is the case, not controlling for these factors could bias coefficient estimates upward, for the parent's SES and school choices could positively correlate to the amount of time that a parent spent engaging in activities related to reading with the child. However, in this paper, several approaches will be taken to address this problem. First, family income will be included in the regressions. Second, by using a school fixed effects model, the school where the child goes to will also be captured.

On the other hand, should there be a negative correlation between participating in these reading activities and the child's reading ability and this remains uncontrolled for in the model, which imposes an omitted variables bias, then coefficient estimates on the relation between family reading activities and student achievement could be driven to zero or negative.

WLS regression results suggest that, among other reading-related activities, having a parent who would play alphabet toys, play word games, read signs out loud, and watch reading programs on television before the child entered formal schooling would carry an effect on the nine year-old child's PIRLS score. The coefficient estimates associated with the first three activities (play alphabet toys, play word games, and read signs verbally) are positive and statistically significant, while engagement in the last activity (watch reading programs on television) is associated with a negative and statistically significant estimate. For instance, on average, having a parent who participated in all four activities is predicted to increase the individual's PIRLS score by 0.97 points, which translates into an effect size of

[4] As suggested by Tiebout's model on public good provision (Tiebout (1956)).

0.1, holding other attributes constant. Results obtained from running a quantile regression become less clear-cut. The only group that resembles what is found in the WLS regressions is composed of individuals belonging to the 25th or 50th quantile of the conditional score distribution.

The paper is organized as the following. Section 3.2 introduces the data set and the regression framework used in the study. Basic results are recorded in section 3.3, while additional tests are performed in section 3.4. Lastly, concluding remarks and suggestions on future work are discussed in section 3.5.

3.2 Data and Empirical Framework

The data set used is the 2001 version of the Progress in International Reading Literacy Study (PIRLS)[5]. Directed by the International Study Center at Boston College in the United States, it aims at studying the reading literacy level of nine year-olds[6] across 35 countries. In each country, 150 schools were selected using a probability-proportional-to-size method. Among these 150 schools, up to two fourth grade classes are randomly selected to participate in the test. Although not all countries adopt English as (one of) their official language(s), the PIRLS committee hires professional translators to ensure that testing materials are comparable across countries[7]. So far, two cohorts of PIRLS have been administered. The first cohort was introduced in 2001, while a newer cohort was conducted in 2006. With respect to the variety of material tested, PIRLS contains two broad types of questions: literary and informational. The first item serves the purpose of testing a student's ability to understand literal passages, while the second asks the test-taker to acquire and use information given in the text to draw inferences [8].

There are four major survey components in PIRLS: questionnaires given to students, parents, teachers, and principals. The student questionnaire mainly asks the student's background and his or her perception of reading. That is, his or her birthday,

[5]Summary statistics of the sample used can found in Table 3.8.
[6]Which corresponds to fourth graders in most of the countries included.
[7]Countries and their respective language(s) of testing can be found in Table 3.7.
[8]See the Appendix for a detailed description on how the PIRLS test booklet is constructed.

gender, and immigrant status, whether he or she likes to read, how much time he or she spends reading both inside and outside of class, and whether he or she believes reading is important for the future. In the parent survey, a parent is asked to provide information on reading resources available at home, such as the number of books or computer availability, which are tools that could induce students to read more outside of class. In addition, there are questions focusing on activities such as how often the parent spent time playing alphabet toys when his or her child was at a younger age. The teacher questionnaire aims at collecting information on the teacher's background (such as educational attainment, gender, teaching experience, and whether he or she holds a teaching certificate), as well as methods used in the classroom for instructional purposes. For instance, information on frequency of reading aloud and student participation in group activities are collected. Lastly, in the principal survey, principals are asked to report on how safe the school is. For example, whether there are theft problems on campus. In addition, principals give their estimate on their students' reading ability prior to starting first grade. In other words, they provide the percentage of students who knew their alphabet or who possessed the ability to write a few words when students initially entered their first grade education. Other general background information, such as the availability of a library on campus or a public library, are also recorded throughout the four types of questionnaires.

The empirical framework adopted is represented as the following:

$$score_{ist} = \alpha + X'_{ist}\beta + P'_{ist}\gamma + \eta score_{st-3} + S'_{st}\lambda + \alpha_j + (\epsilon_{ijt} + \nu_s) \quad (3.1)$$

In the above equation, the outcome variable, $score_{ist}$ corresponds to student i's PIRLS test score[9] achieved in year t (2001) as a fourth grader who studies in school s. In addition, X, P, and S represent regressors signifying student characteristics, family background, and aspects of the school, respectively. The error term can be decomposed into two parts: ϵ_{ijt}, j being country of residence, which represents other unobserved characteristics that could contribute to $score_{ist}$, and ν_s, which denotes the effect of being grouped by a

[9]In particular, this score refers to the *asastdr* variable recorded in PIRLS, reflecting the standardized raw score attained in all items tested, with the mean for each block being 50 and a standard deviation of 10.

particular school or class on the individual's test score. More specifically, given that there can be multiple observations within the same school, the *cluster* function in *Stata* is used (as represented by ν_s), with s being the school that student i goes to.

Since it is unlikely that participation in PIRLS is purely random[10], a weighted least square (WLS) approach is used to evaluate the relationship between various regressors and $score_{ist}$. In particular, the weight used is the *senwgt* variable given in the PIRLS data set, which is the inverse of the probability of a student who is from a specific class of a particular school s in country j to have participated in the test. The set-up of this weight allows each country in the sample to contribute equally, which is appropriate for an international comparison study.

To capture individual characteristics, X_{ist} includes the student's gender, month and year of birth, and whether the child is an immigrant (that is, being born in a country different from the student's current country of residence j). Also included in X_{ist} is a set of dummy variables signifying student i's language(s) spoken. In particular, since there are countries where more than one languages might be used on a regular basis[11], it is important to capture one's language used when he or she was younger. In the student questionnaire of PIRLS, questions such as whether i learned to speak the specific language(s) used in the country where the PIRLS test took place when he or she was little are asked. Also included is the frequency of usage of language of test at home. By incorporating answers to these questions, dummy variables denoting whether i learned to speak the country of residence's language when he or she was younger, as well as whether that language is used at home, are created.

[10] In PIRLS, a two-stage stratification of the sample is used, with the first being a sample of school and the second being a sample of classroom(s) from the same school. Nevertheless, since it is possible that there exists other reasons for a student being absent on the day of test (such as being sick), there might be bias in who actually participated in the test.

[11] For example, Canada adopts both English and French as its official languages. In addition, even if there is only one official language in a country, a child who has an immigrant background might not necessarily use that language for communication purposes at home. For instance, although English is the main language spoken in the United Kingdom, a student whose family immigrated from China might speak Chinese, instead of English, at home. To capture this possibility, answers to the question *How often do you speak (language of test) at home* is used to form a dummy variable set to unity for individuals who answered *Always or Almost Always* and *Sometimes*. This variable is then included as one of the regressors in all of the regression results presented in this paper.

In addition, background information on i's parent (or primary caregiver) is represented by P_{ist}, which lists the parent's educational attainment, reading habit, family income, and whether the parent was born in i's current country of residence[12]. Moreover, P_{ist} also measures the amount of exposure to reading-related activities that the parent participated with i when i was younger (i.e., before school-entry age), which are the main variables of interest in this paper. Dummy variables signifying the correspondences to questions related to reading preparation[13] are further incorporated in P_{ist}. The main hypothesis is that γ would be non-zero. That is, parent participation in activities related to reading with the child before the child started attending school would carry an effect on how well the child performs in the PIRLS test.

As suggested in Betts, Zau, and Rice (2003) and related literature, measuring test score without capturing the student's prior performance might be an incomplete method to study how a policy affects the student's outcome. To take this concern into consideration, information given in the principal survey, which include the principal's estimate on the percent of students who could recognize most of the letters of the alphabet, as well as read and/or write some words or sentences, when they began their first year of formal schooling, will be used to proxy for the student's prior reading knowledge before taking the PIRLS test. These measures are then included in $score_{st-3}$ in the above equation. Usage of this variable is associated with a weakness that it is not a student-specific measure[14].

There are various issues that need to be taken into account before valid conclusions

[12]Although only one parent is asked to fill out the parent survey in PIRLS, both of the parents' (or step-parent(s)'s) education achievement and immigrant status are asked.

[13]These questions include *Before your child began school, how often did you or someone else in your home read books*, *tell stories*, *sing songs*, *play with alphabet toys*, *do reading activities on the computer*, *play word games*, *write letters or words*, *read aloud signs and labels*, and *watch television programs that teach reading, like Sesame Street, with him or her?*. Answers to these questions are represented by three categories: *Often*, *Sometimes*, and *Never or Almost Never*. For those who answered *Often* or *Sometimes* to the above questions, the dummy variable representing each of these reading-related questions will be assigned a value of one.

[14]In the parent survey, parent was also asked to rate the student's reading ability prior to studying first grade. Although answers to these questions would provide a more accurate measure at an individual level, they could also introduce more bias, since the parent might have an intention to over-rate his or her child's ability. In addition, there are more missing values to answers provided in the parent survey compared to those given in the principal survey. Nevertheless, a separate set of regression is run using answers to the parent survey. Since the main results are similar to those obtained using answers given in the principal survey, they are not reported.

can be drawn from this study. One possible complication is the possible correlation between learning atmosphere and student performance. That is, a better learning environment might induce a student to achieve a higher score in PIRLS. If one believes that problems of various types of crime to be a good indicator of school environment, then information on school safety, such as whether theft is a problem at school, which are recorded in the principal survey, can be used. Characteristics of the school, including the percent of students who came from an economically disadvantaged background, and whether the principal perceives theft, vandalism, profanity, and cheating as problems in the school, which are aspects that could affect the student's performance in PIRLS, are captured in S_{st}.

Another possible problem lies in the nature of the PIRLS study. Provided that PIRLS is given to 35 different countries[15], there might be unobservable country characteristics that affect how well the student fares in the test. For instance, cultural characteristics of a particular country could induce a student to work harder in school. By being unable to capture such effect, coefficient estimates may become biased. In order to address this problem, a country fixed effects framework, given as α_j, where j represents each student's country of residence, is incorporated into equation (3.1).

3.3 Regression Results

3.3.1 Findings from WLS

To study whether parent participation in activities related to reading would carry any effect on the child's performance in PIRLS, baseline WLS regressions using the standardized raw score in all items as the outcome variable are run. Related results are reported

[15]Since data from parent survey are unavailable for Morocco and the United States, they are omitted from the sample used in this paper. As a result, only 33 countries are included. They are Argentina, Belize, Bulgaria, Canada, Colombia, Cyprus, Czech Republic, England, France, Germany, Greece, Hong Kong, SAR, Hungary, Iceland, Islamic Republic of Iran, Israel, Italy, Kuwait, Latvia, Lithuania, Republic of Macedonia, Republic of Moldova, the Netherlands, New Zealand, Norway, Romania, Russian Federation, Scotland, Singapore, Slovak Republic, Slovenia, Sweden, and Turkey.

in Table 3.1[16].

One of the interesting observations drawn from Table 3.1 is that, on average, a male student (assigned a value of 1 in the dummy variable *Boy*) is predicted to perform worse compared to a female student. Also, there does not seem to be a statistically significant correlation between whether the student speaks the language that the PIRLS test was conducted in when he or she was little (set to 1 in the dummy variable *Speak Language of Test when Little* if the answer is yes) and his or her PIRLS score. However, if the student uses the language of test at home, his or her test score is predicted to be about 3 points higher than someone who does not do that on a regular basis[17], which suggests that, in order to perform well in a language test, it is more important to practice the language on a regular basis, rather than simply being exposed to it when an individual was younger.

The effects of a parent's participation in reading-related activities with the child when the child was at a preschool age on the PIRLS reading score are represented in column (3) of Table 3.1. More specifically, these activities include reading books (*Read Book*), telling stories (*Tell Story*), singing songs (*Sing Song*), playing alphabet toys (*Play Alphabet Toy*), using the computer to perform reading-related activities (*Computer*), playing word games (*Play Word Game*), writing letters or words (*Write Letters or Words*), reading signs and label out loud (*Read Signs and Labels Aloud*), and watching programs related to reading on television (*Watch TV Reading Programs*) together with the child. According to the regression results, four of these activities are predicted to influence an individual's reading test score: playing alphabet toy, playing word game, reading signs and labels out loud, and watching reading programs on television[18]. Among these, the first three

[16] In all of the regression results presented heretofore, in addition to variables related to the parent's participation in reading-related activities with the child when he or she was at a preschool age, family background variables, namely, the parent's frequency of reading (by themselves) per week (less than an hour, between 1 to 5 hours, between 6 to 10 hours, or more than 10 hours), immigrant status (for both parents), educational attainment (for both parents), and annual family income, are included in the regression model as well.

[17] The relevant dummy variable created, *Speak Language of Test at Home*, is based on the student's response to the question *How often do you speak (language of test) at home?*. This dummy variable is set to equal 1 for an individual who reports speaking the language *Always or Almost Always* or *Sometimes*, and 0 for someone who reports *Never*.

[18] It is also possible that having a parent, especially a mother, who works only part-time to have an effect on the child's performance on PIRLS. By working part-time only, the mother would be able to devote more time taking care of the child, which could induce intellectual stimulation, thus enabling the child

coefficients carry a positive sign, while the last is associated with a negative sign. The signs of the first three activities, playing alphabet toys, playing word games, and reading signs aloud, seem to agree with the general perception that, on average, exposure to words and letters at an early age is related to performing better in a reading test, for it might induce the student to become interested in learning how to read and/or help in building up his or her knowledge in a particular language.

However, despite the fact that dummy variables indicating the amount of television that a student watches per day (*More than 5 Hours*, *Between 3 and 5 Hours*, *Between 1 and 3 hours*, and *Up to an Hour*) have been incorporated in the regression model, the sign associated with the watching reading programs on television estimate is still negative, which is a puzzling finding. Although it might be true that watching too much television could hinder the learning development of students (such as by reducing time spent on studying), being exposed to programs related to reading could, nevertheless, still induce a deeper understanding on language. One possible explanation for the negative coefficient associated with the *Watch TV Reading Programs* variable could be due to other unobserved adverse effect of watching television on student learning. For example, if watching television is negatively correlated with the student's ability to learn how to write (which is an essential skill in taking a written test such as PIRLS), then not incorporating this writing ability measure in the regression could induce the coefficient estimate associated with watching television to be negative[19]. Another possibility is that parent feels obligated to watch

to perform better in school. To test this possibility, a dummy variable signifying the mother's working status (set to unity for a mother who works full-time, zero otherwise) is created and incorporated into the regression model. According to the theory that having a mother who can devote more time with the child would enable the child to learn more, the coefficient is expected to be negative. However, regression results indicate that it is positive and statistically significant instead, showing a positive correlation between having a mother who works full-time and the child's PIRLS score. A possible explanation is that, having a mother who works full-time may mean there are more resources available at home, which help in the child's learning process. With respect to the main variables of interest, namely, activities related to reading that the parent participated in, they are similar to those reported in column (3) of Table 3.1, except that the coefficient associated with *Play Alphabet Toy* remains positive, but becomes statistically insignificant.

[19] The negative sign associated with *Watch TV Reading Programs* may be due to how the dummy variable is defined (see footnote 13). As a robustness check, in a separate set of regressions, the associated reading-related dummies are redefined according to their original classification scheme given in the PIRLS questionnaire. As a result, 3 sets of dummy variables, *Often*, *Sometimes*, and *Never*, with the last category being the omitted group, are created for each of the reading-related activities listed in Table 3.1. However, the coefficient estimates for the *Often* and *Sometimes* dummies associated with the *Watch TV Reading*

television with the child only if he or she has problems reading[20].

Upon running a correlation matrix among the parent contribution variables (listed in Table 3.2), it turns out that some variables have a correlation higher than 0.70. For instance, the correlation between *Read Book* and *Tell Story* is 0.88. If these are driven by the parent's preference to engage in activities related to reading, then inclusion of both variables in the regression model could reduce the significance of (one of the) coefficient estimate(s). To further test whether the negative coefficient estimate for the *Watch TV Reading Programs* is due to the high correlation between variables, another set of regressions including only a few of the reading-related variables (*Read Book*, *Play Alphabet Toy*, *Computer*, *Play Word Games*, and *Watch TV Reading Programs*) is run[21]. It turns out that the coefficient estimate associated with the *Watch TV Reading Programs* remains negative and statistically significant, indicating that a high correlation between variables does not help explain why watching reading programs on television is predicted to carry an adverse effect on PIRLS score.

As discussed in the introduction, the choice of which school to attend might be correlated to the family's SES. To address this issue, columns (4) and (5) of Table 3.1 report regression results obtained under a school fixed effects framework. In general, those in column (4) are similar to that reported in column (3), except for the coefficient estimates on the *Play Alphabet Toy* and *Read Signs and Labels Aloud* terms. More specifically, although positive, they both have become statistically insignificant[22]. To test whether this result is driven by the sample selection criteria, that is, using a sample that only includes observations with non-missing responses to all of the regressors included, an imputed miss-

Programs remain negative and statistically significant.

[20]This scenario may happen more often if the parent is an immigrant himself or herself. To test this possibility, regression results obtained from including only immigrants in the sample can be found in Table 3.5. According to Table 3.5, none of the parental input variables are statistically significant. These findings will be discussed in the section 3.4.2.

[21]Judging from their correlation with the *Read Book* variable, those with a correlation close to or higher than 0.80, namely, *Tell Story*, *Sing Song*, *Write Letters and Words*, and *Read Signs and Labels Aloud*, are dropped from the model. Alternative regressions using other combinations of reading-related variables are also run. Given that regression results obtained are highly similar to those reported in Table 3.1, they are, therefore, not reported.

[22]The coefficient on *Play Alphabet Toy* is, however, still statistically significant at a 10% level.

ing value methodology was used[23] along with the school fixed effects framework. As shown in column (5) of Table 3.1, the signs on the coefficient estimates of the variables of interest are very similar to those reported in column (3), except for the *Sing Song* term, which is negative and statistically significant[24]. One possible explanation is that, although singing songs with the child could enhance the child's listening and speaking skills, it may not improve his or her writing skill, which is the main method of testing in PIRLS. In addition, if it is the case that a parent from a less equipped household tends to sing more songs with the child, for there are less resources available at home, then participation in this activity could be correlated to some unobserved family background measures. As a result, devoting time singing songs with a child might carry a negative association with the child's PIRLS score.

To further study the effect of parental input in activities related to reading on a child's reading score, suppose a comparison is drawn between two students who are identical in other attributes except that the first student's parent was involved in playing alphabet toys, playing word games, reading signs and labels out loud, and watching reading programs on television before the student began formal schooling. Using coefficient estimates obtained in column (3) of Table 3.1, the former student is predicted to score 0.97 points higher in PIRLS. Moreover, the associated effect size is calculated to be 0.1[25]. Furthermore, had the parent engaged in the first three activities except for watching reading programs on television with the child, the effect size would have been doubled (0.2 instead of 0.1). This finding suggests that parent involvement in playing alphabet toys, playing word games, combined with reading signs and labels aloud would contribute positively to a student's performance in the reading test.

[23] That is, another data set, which includes all possible observations including those with missing values, is created. Responses with a missing value are first assigned a value of zero and then indicated with a dummy variable signifying that imputed values are used for these observations. By doing so, sample size greatly increased from 19950 to 31740.

[24] However, the coefficient estimate associated with this term is marginally significant at a 10% level (column (3)) as well.

[25] Note that the outcome measure, *asastdr*, is the raw score on all items given in the PIRLS test, standardized to be mean 50 with a standard deviation of 10. Effect size is calculated as the mean difference in PIRLS score between two individuals, one exposed to the "treatment" of having a parent participated in reading-related activities, while the other was not, divided by the standard deviation, or 10.

3.3.2 Quantile Regression Results

The results presented in Table 3.1 represent an average effect of a parent's engagement in reading-related exercises on student outcome. To further investigate whether their significance is conditional upon how well a student performs in PIRLS, quantile regressions, which divide students according to where they stand in the conditional PIRLS score distribution (10th, 25th, 50th, 75th, or 90th quantile), are run. Results are presented in Table 3.3.

Similar to results found in Table 3.1, on average, a boy is predicted to perform worse than a girl in the test. In addition, there is a strong correlation between speaking the language used in conducting the test at home on a regular basis and the student's test performance. This correlation seems to be the strongest for someone belonging to the lower 10th quantile of the conditional distribution: compared to someone who does not use the language adapted in the written test at home, a student who speaks the language regularly would score 3.35 points higher in PIRLS.

In general, parent involvement in reading books, telling stories, singing songs, using a computer for reading purposes, or writing words or letters when the student was at a preschool age is predicted to carry no effect on the student's performance in PIRLS, for the coefficient estimates associated with these variables are statistically insignificant across all quantiles. Overall, a parent's involvement in reading-related activities is predicted to affect the test score achieved by a student at the lower quartile of the conditional distribution the most. For example, it is only at the 25th quantile that the coefficient estimates of the *Play Alphabet Toy*, *Play Word Game*, and *Read Signs and Labels Aloud* are all found to be positive and statistically significant[26]. Precisely, holding other attributes constant, for an individual whose parent has spent time playing alphabet toys, playing word games, reading signs and labels out loud, and watching reading programs on television in his or her preschool years, compared with another student whose parent has not gotten involved in these activities but scored at the same lower quartile of the conditional distribution,

[26]The coefficient estimate of *Play Word Game* is also positive and statistically significant for an individual at the 50th quantile of the conditional score distribution, meaning that, for someone belonging to the median quantile, having a parent who would spend time playing word games with him or her is predicted to increase his or her PIRLS score by 0.49 points, holding other variables constant.

the former student is predicted to score 1.4 points higher in PIRLS. On the other hand, a parent's participation in reading-related activities, except for the adverse effect of watching television programs related to reading, is seen as carrying no effect on the student's test score, should he or she score at the 75th or the 90th quantile of the conditional distribution.

Another observation is that an individual at the lower 10th quantile of the conditional score distribution does not seem to be affected by the adverse effect of having a parent who would watch reading programs with him or her before he or she entered school. As evident from the last row in Table 3.3 (*Watch TV Reading Programs*), the associated coefficient estimate is negative and statistically significant for all quantiles except for the 10th. Additionally, the magnitude of this adverse effect decreases as the student scores higher in PIRLS, holding other variables fixed. For instance, having a parent who spent time watching television programs related to reading, holding other characteristics constant, is predicted to decrease someone at the 25th quantile of the conditional distribution's test score by 1.02 points, while this effect decreases to 0.64 points for another student at the 90th quantile of the same distribution.

3.4 Additional Tests

3.4.1 The Effect of Teacher Characteristics on Student Outcome

One of the main interests in the literature is to study whether there is a correlation between teacher quality and the student's education outcome, such as test score. However, researchers have not found a unanimous consensus on whether teacher characteristics contribute to how well the student fares in a test. In addition, there might be a correlation between teacher quality and family background, which could affect a child's test performance. More specifically, coming from a family with a high SES might enable a parent to select a more qualified teacher for the child, such as through having the choice of going to a private or public school, which is unaccessible for someone from a low SES family. As a result, the child might be able to score higher in a test, since being taught by a more qualified teacher might induce him or her to learn more. To investigate this possibility, various aspects of the teacher are added to the regression model presented in equation (3.1). These aspects include the teacher's gender, teaching experience, and whether he or she holds a teaching certificate. Results are summarized in Table 3.4[27].

As shown in column (3) of Table 3.4, there is a negative correlation between test score and being taught by a male teacher, for the coefficient estimate associated with *Male* (a dummy variable set to one for a male teacher) is negative. Also, teaching experience (*Experience*) does not seem to matter, while, on average, having a teacher who holds a teaching certificate (represented by *Certification*, a dummy variable set to 1 for a teacher holding a credential) is predicted to positively affect the student's PIRLS score.

Although various teacher characteristics seem to affect the individual's test score, the signs of the coefficients of the main variables of interest, or those reflecting reading activities that the parent engaged in before the individual formally attended school, do not seem to differ from those found in the basic WLS regression (column (3) of Table 3.1). In particular, whether the parent has spent time playing alphabet toys, playing word games, and reading signs and labels out loud are predicted to positively affect student outcome, while there seems to be a negative relationship between watching reading programs on television and the student's average performance in PIRLS.

As discussed before, including only data points that contain no missing values to all of the regressors may be correlated with selection bias. To test whether results obtained so far are due to the inclusion of a selected group of individuals in the sample, an imputed missing value methodology is utilized[28]. Results are presented in column (4) of Table 3.4. Comparing columns (3) and (4), one can see that there is no major deviation in the signs of the coefficient estimates. That is, in column (4), the signs for *Speak Language of Test at Home*, *Play Alphabet Toy*, *Play Word Game*, *Read Signs and Labels Aloud*, and *Certification* remain positive, while those associated with *Boy*, *Watch TV Reading Programs*, and *Male* teacher are negative. In other words, results acquired under the use of imputed missing

[27]To test whether the effect of teaching experience on the student's PIRLS score would depend on the level of experience that the specific teacher has, a separate regression using a set of dummy variables denoting a teacher with *less than or equal to 3, between 4 and 6, and more than 6 years of experience* is run. Since results highly resemble those reported in Table 3.4, they are not reproduced here.

[28]Again, this refers to first replacing missing values with zero numerically, then using dummy variables to represent these imputed responses in the data set.

values are very similar to those obtained through including only observations with no missing values in the sample. Focusing on the main variables of interest, it still seems to be safe to conclude that the correlation between playing alphabet toys, playing word games, and reading signs and labels out loud and a student's performance in PIRLS is positive. On the other hand, a parent's involvement in watching reading programs on television is predicted to negatively affect the student's PIRLS score on average.

3.4.2 Testing Whether Results Remain Robust for Students with an Immigrant Background

Another extension to the regression framework used so far is to test whether the effect of a parent's participation in reading-related activities on the student's PIRLS performance would persist for a student who is an immigrant. Assuming an immigrant to be fluent in a language different from that used in the host country, the ability to read could become even more crucial for him or her to succeed in the country of destination. As a result, conducting a study on the immigrant sub-sample would be interesting. Related results are given in Table 3.5. For comparison purposes, coefficient estimates obtained from the whole sample (column (3) in Table 3.1) and when teacher characteristics are added (column (3) in Table 3.4) are reproduced in columns (1) and (2).

Similar to when the whole sample and when teacher quality are added, being a boy immigrant is negatively correlated to one's performance in PIRLS. In addition, speaking the host country's language at home on a regular basis is predicted to increase one's test score by 3.7 points on average (column (3) in Table 3.5). However, the major difference between results listed in columns (1) and (2) and those acquired under the use of an immigrant sub-sample is that a parent's involvement in various reading-related activities before an immigrant student attended formal schooling is predicted to have no effect on his or her PIRLS score[29]. One possible explanation is that, for an immigrant student, even if his or her parent did spend time to take part in reading-related activities when he or she

[29]Note that the signs of the *Tell Story*, *Computer*, and *Read Signs and Labels Aloud* in column (3) are the opposite of those reported in columns (1) and (2). However, their coefficient estimates are statistically insignificant.

was at a preschool age, if the language used during that time was that of the country of origin's and this language is different from that used in the country of destination, receiving training when young might not necessarily carry an effect on how well the student performs in PIRLS, a test that might be conducted in another language[30].

As I have previously discussed, one reason for none of the parent participation variables to be statistically significant when the immigrant sub-sample is used could be because of the existence of high correlation between variables used. For example, a parent who would read books with the child may also be more likely to tell stories to the child. To test whether it is the high correlation among variables that produce insignificant estimates in the immigrant sub-sample, Table 3.6 reduces the number of variables relevant to academic activities included in the regression framework. More specifically, only the *Read Book*, *Play Alphabet Toy*, *Computer*, *Play Word Game*, and *Watch TV Reading Programs* are kept in the model[31].

As evident in column (3) of Table 3.6, none of the reading-related parental input variables are statistically significant, a finding similar to that shown in column (3) of Table 3.5, which includes a complete set of activities that develop reading skills in the framework[32]. This may suggest that, although there is a high correlation between different variables describing parent contribution in reading-related activities before the student started first grade, for an immigrant student, parental input in these activities seems to carry no statistically significant effect on his or her average PIRLS score. If this immigrant

[30] To test whether this effect would be due to the omission of observations containing missing values, the imputed missing values methodology is used. Results are summarized in column (4) in Table 3.5. An interesting observation drawn through comparing results obtained from using an immigrant sample with a full response rate on all variables and from a sample that is composed of imputed values is the change in signs for some of the coefficient estimates. In particular, signs for the *Speak Language of Test when Little*, *Tell Story*, *Write Letters or Words*, and *Watch TV Reading Program* all become positive in the sample that includes imputed missing values (column (4)). However, these estimates are, again, statistically insignificant, meaning that they are predicted to carry no effect on an immigrant student's PIRLS score on average.

[31] See footnote 21 for an explanation on why only these variables are kept in the model. These results remain robust when alternative regressions using other combinations of reading-related variables are run. That is, none of the associated coefficient estimates are statistically significant and therefore not reported.

[32] Although the magnitude of the coefficient estimates tends to be lower (higher) for *Read Book*, *Play Alphabet Toy*, *Computer*, and *Play Word Game* (*Watch TV Reading Programs*) when only these variables are included in the empirical model, compared to those found when more reading-related regressors are used (column (3) of Table 3.5).

student got exposed to reading in a language that is different from that used to conduct the PIRLS test, then it is understandable to not observe a link between a parent's engagement in reading-related activities at the child's young age and the child's subsequent PIRLS performance[33].

However, notice how the sample size becomes a lot smaller when only immigrants are included in the sample (containing 2,890 observations, compared to the 19,950 observations in the full-sample). Having a smaller sample size implies that standard errors would also become larger, thus might reduce the significance of many variables. Therefore, it may be this reduction in sample size that drives the relationship between parental input and an immigrant child's PIRLS test score to become insignificant.

3.5 Conclusion

Although many researchers believe in the importance of family background on student outcome, not many studies have decomposed family background further to study which part of it would contribute the most to the well-being of students. Using data and questionnaires collected in the 2001 Progress in International Reading Literacy Study (PIRLS), which is a reading test given to nine year-old students across 35 countries, this paper studies the contribution of having a parent who would participate in various reading-related activities before the child reached a school entry age to his or her performance in the test. More specifically, reading-related activities are defined in nine categories: *reading books*, *telling stories*, *singing songs*, *playing alphabet toys*, *using the computer for reading purposes*, *playing word games*, *writing letters or words*, *reading signs and labels aloud*, and *watching programs related to reading on television*. Among these activities, having a parent who would play alphabet toys, play word games, or read signs and labels verbally is predicted to positively affect a student's PIRLS performance, while watching reading programs on television seem to adversely affects one's test score on average. More precisely, the effect size associated with having a parent participating in these four exercises

[33] Regression results obtained under the use of an imputed missing values sample is listed in column (4) of Table 3.6. Again, none of the parental input variables is statistically significant.

is calculated to be 0.1. The magnitude of this effect size would have doubled should a parent participate in the first three of the activities listed above (playing alphabet toys, playing word games, and reading signs and labels aloud), but not watching reading programs on television, before the child entered school.

In addition to being able to investigate what type of reading-related exercises the parent engaged in during the child's preschool years, PIRLS also provides information on several teacher characteristics, such as gender, education, experience, and whether the teacher holds a teaching certificate. Upon inclusion of these independent variables, the main results found in this paper, namely, that parent participation in reading-related activities is predicted to positively affect the student's performance in PIRLS on average, remains robust.

Results obtained through the use of a quantile regression framework show that the effect of these activities (reading books, telling stories, singing songs, playing alphabet toys, using the computer for reading purposes, playing word games, writing letters or words, reading signs and labels out loud, and watching reading programs on television) on student outcome differs, with most of the effects focused on students below the 50th quantile of the conditional test score distribution.

Based on results reported in this paper, one possible implication drawn is the importance of parent involvement during the early years of a child, especially in terms of developing the child's reading skills. Additionally, usage of the relevant language on a regular basis (such as speaking it at home) also contributes to how well the student fares in the test (The coefficient estimate of the *Speak Language of Test at Home* term is positive and statistically significant, regardless of the sample and the regression framework used.). Although parental input is predicted to carry no effect on the score attained by an immigrant student, constant usage of the language at home is still a significant factor in affecting his or her performance in the reading test.

One possible extension to this project is to study whether parental contribution would affect the outcome of the student across different subjects, such as mathematics and science. The biggest obstacle in achieving this task is data availability. Some possible candidates include the Trends in International Mathematics and Science Study (TIMSS), an international test that takes place every four years, which is composed of questions in the fields of mathematics and science for fourth and eighth graders, as well as the Program for International Student Assessment (PISA), a test given to fifteen year-olds across different countries, aiming to measure their ability in reading, mathematics, and science. Should background variables included in these studies be as rich as those given in PIRLS, conducting research on whether similar effects of parental input on a student's test score continue to be observed in another subject area would be possible. Another potential future research extension is to perform the same analysis on the 2006 cohort of PIRLS, in order to observe whether the predicted positive effect of playing alphabet toys, playing word games, reading signs and labels out loud, and the negative effect of watching reading programs shown on television on the individual's average PIRLS score persist across the 2001 and 2006 cohorts.

Table 3.1: The Effect of Parent Participation in Reading-Related Activities in Preschool Years on the Child's Average PIRLS Performance

	(1)	(2)	(3)	(4)	(5)†
Boy	-1.9778*	-1.9144*	-1.9144*	-1.7952*	-1.9493*
	(0.1536)	(0.1521)	(0.1683)	(0.1890)	(0.1353)
Speak Language of Test when Little	0.0061	0.0548	0.0548	0.3215	0.5690
	(0.2792)	(0.3913)	(0.4115)	(0.4850)	(0.3302)
Speak Language of Test at Home	2.1701*	2.9730*	2.9730*	2.7747*	2.3882*
	(0.4681)	(0.4709)	(0.4443)	(0.5193)	(0.3842)
Read Book	0.1135	0.4016	0.4016	0.5250	0.4090
	(0.3812)	(0.3821)	(0.3845)	(0.4292)	(0.3045)
Tell Story	0.2539	0.1700	0.1700	0.1798	0.2156
	(0.3325)	(0.3367)	(0.3516)	(0.4108)	(0.2837)
Sing Song	-0.4634	-0.4238	-0.4238	-0.2796	-0.4156*
	(0.2525)	(0.2499)	(0.2257)	(0.2947)	(0.1953)
Play Alphabet Toy	0.3244	0.6072*	0.6072*	0.5307	0.4536*
	(0.2271)	(0.2379)	(0.2555)	(0.2829)	(0.2050)
Computer	-0.0640	-0.1288	-0.1288	-0.4482	-0.2749
	(0.1990)	(0.2012)	(0.2331)	(0.2656)	(0.1902)
Play Word Game	0.7151*	0.6417*	0.6417*	0.5772*	0.7169*
	(0.2217)	(0.2257)	(0.2133)	(0.2590)	(0.1863)
Write Letters or Words	-0.4701	-0.4388	-0.4388	0.0306	-0.1565
	(0.2936)	(0.2903)	(0.2897)	(0.3251)	(0.2269)
Read Signs and Labels Aloud	0.4930*	0.6593*	0.6593*	0.3961	0.3533*
	(0.2448)	(0.2457)	(0.2245)	(0.2709)	(0.1813)
Watch TV Reading Programs	-0.3370	-0.9385*	-0.9385*	-0.9728*	-0.6430*
	(0.1767)	(0.2117)	(0.1999)	(0.2523)	(0.1692)

Table 3.1: The Effect of Parent Participation in Reading-Related Activities in Preschool Years on the Child's Average PIRLS Performance, continued

	(1)	(2)	(3)	(4)	(5)†
F(Country FE Dummies Jointly Zero)		23.48	0.79	5.07	0.80
p-value		[0.0000]	[0.7905]	[0.0000]	[0.7815]
Country Fixed Effects	No	Yes	Yes	Yes	Yes
School Cluster	No	No	Yes	Yes	Yes
School Fixed Effects	No	No	No	Yes	Yes
N	19950	19950	19950	19950	31740

[1] The † in column (5) denotes that the sample used contains imputed missing values, which serves the purpose of testing whether the insignificance in the coefficients of interest is due to reduction in sample size. These imputed values mean that, unlike results shown in columns (1) through (4), which use a sample that only contains a 100% response rate on variables included, those listed in column (5) assigns zero to any response that contains a missing value and subsequently uses dummy variables to represent these observations with missing values. As a result, the number of observations greatly increased from 19950 to 31740.

[2] * signifies 5% statistical significance.

[3] For a detailed explanation of variables presented, see text.

Table 3.2: Correlation Matrix between Reading-Related Activities that a Parent Participated in Before the Student Began Formal Schooling

	Read Book	*Tell Story*	*Sing Song*	*Play Alphabet Toy*	*Computer*	*Play Word Game*	*Write Letters, Words*	*Read Signs Aloud*	*Watch Reading Programs*
Read Book	1.00								
Tell Story	0.88	1.00							
Sing Song	0.81	0.82	1.00						
Play Alphabet Toy	0.76	0.77	0.69	1.00					
Computer	0.25	0.22	0.23	0.24	1.00				
Play Word Game	0.74	0.73	0.68	0.73	0.27	1.00			
Write Letters and Words	0.85	0.83	0.78	0.75	0.24	0.73	1.00		
Read Signs Aloud	0.79	0.78	0.72	0.70	0.25	0.70	0.79	1.00	
Watch Reading Programs	0.57	0.55	0.55	0.49	0.26	0.49	0.55	0.53	1.00

[1] Variables presented on the vertical and horizontal axes are reading-related activities that a parent participated in before the student entered school. For a detailed description of each variable's meaning, see text.

Table 3.3: The Effect of Parent Participation in Reading-Related Activities in Preschool Years on the Child's Average PIRLS Performance, Quantile Regression Results

Quantile	10th (1)	25th (2)	50th (3)	75th (4)	90th (5)
Boy	-2.4709*	-2.4730*	-1.7444*	-1.4085*	-1.0318*
	(0.2832)	(0.1307)	(0.1555)	(0.1264)	(0.1416)
Speak Language of Test when Little	1.0103	0.8286	0.1875	0.3320	0.7594*
	(0.7290)	(0.5001)	(0.5551)	(0.2780)	(0.3841)
Speak Language of Test at Home	3.3529*	3.1406*	3.0171*	2.4373*	1.7459*
	(0.9330)	(0.3869)	(0.5140)	(0.4907)	(0.5219)
Read Book	0.5066	0.3612	0.4435	0.5100	-0.2389
	(0.7795)	(0.5055)	(0.3450)	(0.3427)	(0.2910)
Tell Story	0.8720	0.2848	-0.0083	0.1610	0.3393
	(0.5409)	(0.4179)	(0.3342)	(0.3015)	(0.2697)
Sing Song	-0.4325	-0.5633	-0.3668	-0.2491	-0.0864
	(0.3487)	(0.4447)	(0.2217)	(0.2252)	(0.1721)
Play Alphabet Toy	0.4925	0.6177*	0.4813	0.2728	0.2817
	(0.3812)	(0.3029)	(0.3800)	(0.2536)	(0.2271)
Computer	-0.4142	-0.3370	-0.2390	-0.1961	-0.0937
	(0.5167)	(0.3421)	(0.1980)	(0.1901)	(0.1656)
Play Word Game	0.6582	0.9762*	0.4931*	0.3401	0.4493
	(0.4117)	(0.2676)	(0.2150)	(0.1758)	(0.2355)
Write Letters or Words	-0.4439	-0.3089	-0.1548	-0.1404	0.0548
	(0.7178)	(0.2727)	(0.2509)	(0.2404)	(0.2671)
Read Signs and Labels Aloud	0.6934	0.8105*	0.4639	0.3667	0.1741
	(0.4645)	(0.4108)	(0.2844)	(0.2016)	(0.2338)
Watch TV Reading Programs	-0.6414	-1.0152*	-0.9368*	-0.7930*	-0.6446*
	(0.3375)	(0.3572)	(0.2387)	(0.1955)	(0.1281)

Table 3.3: The Effect of Parent Participation in Reading-Related Activities in Preschool Years on the Child's Average PIRLS Performance, Quantile Regression Results, continued

Quantile	10th	25th	50th	75th	90th
	(1)	(2)	(3)	(4)	(5)
F(All Country FE Dummies Jointly Zero)	37.42	79.40	61.12	65.58	77.06
p-value	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes
N	19950	19950	19950	19950	19950

[1] Instead of allowing a school cluster effect (using the *cluster* command in Stata), a bootstrap procedure that replicated the sample 20 times is used to produce these results.

[2] Regression results based on nine year-old students who participated in the 2001 PIRLS reading test and have no missing values in all of the variables of interest. Other regressors include the month and year of birth, frequency of watching television per day (Which is a set of five dummy variables representing *none*, *less than an hour*, *between 1 to 3 hours*, *between 3 to 5 hours*, and *more than 5 hours* per day. The first group is the omitted category.), immigrant status of the student and his or her parents, parental education, parent's reading habit (Represented by a set of four dummy variables: spending *more than 10 hours*, *between 6 to 10 hours*, *between 1 to 5 hours*, and *less than an hour* per week. The last category is the omitted category.), categorical family income dummies, as well as school background variables, including the percent of students who are from an economically disadvantaged background, percent of first graders who could recognize most of the alphabet, read some words and/or sentences, write some letters, alphabet, and/or words in the same school that individual i goes to, and whether there are problems of cheating, profanity, or vandalism. Notice that all of the school background variables are extracted from the principal survey. Country fixed effects dummy variables are also included in some models. Data sources: student, home (parent), teacher, and school (principal) survey of the 2001 PIRLS. All regressors are weighted using the *senwgt* variable presented in PIRLS, which basically lets each country contribute equally to perform an international comparison study.

[3] * signifies 5% statistical significance.

[4] For a detailed explanation of variables presented, see text.

Table 3.4: The Effect of Parent Participation in Reading-Related Activities in Preschool Years on the Child's Average PIRLS Performance, with Teacher Characteristics as Additional Regressors

	(1)	(2)	(3)	(4)†
Boy	-1.9003*	-1.8272*	-1.8272*	-1.8865*
	(0.1592)	(0.1574)	(0.1673)	(0.1373)
Speak Language of Test when Little	-0.1746	-0.0830	-0.0830	0.5588
	(0.2910)	(0.4062)	(0.4145)	(0.3389)
Speak Language of Test at Home	2.2914*	2.8926*	2.8926*	2.4317*
	(0.4944)	(0.4958)	(0.4703)	(0.3944)
Read Book	0.0198	0.3250	0.3250	0.3404
	(0.3971)	(0.3993)	(0.4036)	(0.3129)
Tell Story	0.1280	0.1666	0.1666	0.2635
	(0.3505)	(0.3528)	(0.3749)	(0.2905)
Sing Song	-0.4181	-0.3688	-0.3688	-0.3157
	(0.2605)	(0.2577)	(0.2337)	(0.2001)
Play Alphabet Toy	0.2553	0.6023*	0.6023*	0.4526*
	(0.2354)	(0.2456)	(0.2592)	(0.2086)
Computer	-0.0333	-0.0568	-0.0568	-0.2685
	(0.2046)	(0.2067)	(0.2413)	(0.1951)
Play Word Game	0.6877*	0.5967*	0.5967*	0.7391*
	(0.2271)	(0.2315)	(0.2246)	(0.1923)
Write Letters or Words	-0.5538	-0.4648	-0.4648	-0.2739
	(0.3023)	(0.2994)	(0.2928)	(0.2238)
Read Signs and Labels Aloud	0.4315	0.6351*	0.6351*	0.4146*
	(0.2525)	(0.2535)	(0.2288)	(0.1837)
Watch TV Reading Programs	-0.1470	-0.9127*	-0.9127*	-0.7149*
	(0.1863)	(0.2173)	(0.2015)	(0.1695)
Teacher Characteristics				
Experience	0.0032	0.0057	0.0057	0.0038
	(0.0050)	(0.0050)	(0.0060)	(0.0055)
Male	-1.1520*	-1.1497*	-1.4967*	-1.3721*
	(0.2213)	(0.2313)	(0.2923)	(0.2462)
Certification	0.6663*	0.8338*	0.8338*	1.0475*
	(0.2591)	(0.3047)	(0.3822)	(0.3263)

Table 3.4: The Effect of Parent Participation in Reading-Related Activities in Preschool Years on the Child's Average PIRLS Performance, with Teacher Characteristics as Additional Regressors, continued

	(1)	(2)	(3)	(4)†
F(All Country FE Jointly Zero)		22.03	0.77	0.69
p-value		[0.0000]	[0.8207]	[0.9020]
Country Fixed Effects	No	Yes	Yes	Yes
School Cluster	No	No	Yes	Yes
N	18624	18624	18624	30680

[1] Regression results based on nine year-old students who participated in the 2001 PIRLS reading test and have no missing values in all of the variables of interest. Other regressors include the month and year of birth, frequency of watching television per day (Which is a set of five dummy variables representing *none*, *less than an hour*, *between 1 to 3 hours*, *between 3 to 5 hours*, and *more than 5 hours* per day. The first group is the omitted category.), immigrant status of the student and his or her parents, parental education, parent's reading habit (Represented by a set of four dummy variables: spending *more than 10 hours*, *between 6 to 10 hours*, *between 1 to 5 hours*, and *less than an hour* per week. The last category is the omitted category.), categorical family income dummies, as well as school background variables, including the percent of students who are from an economically disadvantaged background, percent of first graders who could recognize most of the alphabet, read some words and/or sentences, write some letters, alphabet, and/or words in the same school that individual i goes to, and whether there are problems of cheating, profanity, or vandalism. Notice that all of the school background variables are extracted from the principal survey. Country fixed effects dummy variables are also included in some models. Data sources: student, home (parent), teacher, and school (principal) survey of the 2001 PIRLS. All regressors are weighted using the *senwgt* variable presented in PIRLS, which basically lets each country contribute equally to perform an international comparison study.

[2] The † in column (4) denotes that, unlike results shown in columns (1) through (3), which use a sample that only contains a 100% response rate on variables included, those listed in column (4) assigns zero to any response that contains a missing value and subsequently uses dummy variables to represent these observations with missing values. As a result, the number of observations greatly increased from 18624 to 30680.

[3] * signifies 5% statistical significance.

[4] For a detailed explanation of variables presented, see text.

Table 3.5: The Effect of Parent Participation in Reading-Related Activities in Preschool Years on an Immigrant Child's Average PIRLS Performance

	Whole Sample		Immigrants Only	
		with Teacher Qualities		Imputed Missing Values
	(1)	(2)	(3)	(4)†
Boy	-1.9144*	-1.8272*	-2.1449*	-2.0956*
	(0.1683)	(0.1673)	(0.4093)	(0.3259)
Speak Language of Test when Little	0.0548	-0.0830	-0.4687	0.1444
	(0.4115)	(0.4145)	(0.7629)	(0.5644)
Speak Language of Test at Home	2.9730*	2.8926*	3.6907*	3.4494*
	(0.4443)	(0.4703)	(0.9398)	(0.7395)
Read Book	0.4016	0.3250	1.4292	0.3568
	(0.3845)	(0.4036)	(1.2259)	(0.7689)
Tell Story	0.1700	0.1666	-0.1490	0.6278
	(0.3516)	(0.3749)	(0.9016)	(0.6553)
Sing Song	-0.4238	-0.3688	-0.6785	-0.5102
	(0.2257)	(0.2337)	(0.8165)	(0.6001)
Play Alphabet Toy	0.6072*	0.6023*	0.5676	0.8284
	(0.2555)	(0.2592)	(0.7100)	(0.5483)
Computer	-0.1288	-0.0568	0.1313	0.3406
	(0.2331)	(0.2413)	(0.6130)	(0.4925)
Play Word Game	0.6417*	0.5967*	0.8654	0.4031
	(0.2133)	(0.2246)	(0.5854)	(0.4330)
Write Letters or Words	-0.4388	-0.4648	-0.3360	0.7302
	(0.2897)	(0.2928)	(0.9071)	(0.6866)
Read Signs and Labels Aloud	0.6593*	0.6351*	-0.5802	-0.1226
	(0.2245)	(0.2288)	(0.6402)	(0.4969)
Watch TV Reading Programs	-0.9385*	-0.9127*	-0.5154	0.0321
	(0.1999)	(0.2015)	(0.5714)	(0.4557)
F(All Country FE Jointly Zero)	0.79	0.77	1.08	1.27
p-value	[0.7905]	[0.8207]	[0.3546]	[0.1461]
Country Fixed Effects	Yes	Yes	Yes	Yes
School Cluster	Yes	Yes	Yes	Yes
N	19950	18624	2890	4896

[1] * signifies 5% statistical significance.

Table 3.6: The Effect of Parent Participation in Reading-Related Activities in Preschool Years on an Immigrant Child's Average PIRLS Performance, with Fewer Academic Activities

	(1)	(2)	(3)	(4)†
Boy	-2.2695*	-2.1103*	-2.1103*	-2.0595*
	(0.4604)	(0.4403)	(0.4102)	(0.3285)
Speak Language of Test when Little	0.3801	-0.4353	-0.4353	0.1102
	(0.6495)	(0.7399)	(0.7593)	(0.5527)
Speak Language of Test at Home	2.5537*	3.6598*	3.6598*	3.2991*
	(0.8930)	(0.9412)	(0.9436)	(0.7355)
Read Book	0.9960	0.7836	0.7836	0.0530
	(0.8975)	(0.9193)	(0.9336)	(0.6118)
Play Alphabet Toy	0.4710	0.3266	0.3266	0.6587
	(0.6609)	(0.6868)	(0.6943)	(0.5456)
Computer	0.0746	0.0847	0.0847	0.2627
	(0.5695)	(0.5680)	(0.6073)	(0.4853)
Play Word Game	0.5116	0.6443	0.6443	0.3082
	(0.6162)	(0.6056)	(0.5565)	(0.4127)
Watch TV Reading Programs	-0.5039	-0.7144	-0.7144	-0.2016
	(0.5545)	(0.6163)	(0.5666)	(0.4392)
F(All Country FE Jointly Zero)		5.54	1.07	1.42
p-value		[0.0000]	[0.3719]	[0.0636]
Country Fixed Effects	No	Yes	Yes	Yes
School Cluster	No	No	Yes	Yes
N	2890	2890	2890	4896

[1] Due to the high correlation between parental input variables, some of them are dropped in this set of regression results. Variables dropped include whether the parent participated in telling story, singing song, writing letters or words, and reading signs and labels aloud with the student during his or her preschool years.

[2] The † in column (4) denotes that the sample used contains imputed missing values, which serves the purpose of testing whether the insignificance in the coefficients of interest is due to reduction in sample size. These imputed values mean that, unlike results shown in columns (1) through (3), which use a sample that only contains a 100% response rate on variables included, those listed in column (4) assigns zero to any response that contains a missing value and subsequently uses dummy variables to represent these observations with missing values. As a result, the number of observations greatly increased from 2890 to 4896.

[3] * signifies 5% statistical significance.

[4] For a detailed explanation of variables presented, see text.

Table 3.7: Test Language(s) Used in PIRLS for the 33 Countries Included in the Sample

Country	Language(s) Used
Argentina	Spanish
Belize	English
Bulgaria	Bulgarian
Canada	English or French
Colombia	Spanish
Cyprus	Greek
Czech Republic	Czech
England	English
France	French
Germany	German
Greece	Greek
Hong Kong, SAR	Chinese
Hungary	Hungarian
Iceland	Icelandic
Islamic Republic of Iran	Farsi
Israel	Hebrew or Arabic
Italy	Italian
Kuwait	Arabic
Latvia	Latvian or Russian
Lithuania	Lithuanian
Republic of Macedonia	Macedonian or Albanian
Republic of Moldova	Romanian (Moldovan) or Russian
Netherlands	Dutch
New Zealand	English
Norway	Bokmal or Nynorsk
Romania	Romanian or Hungarian
Russian Federation	Russian
Scotland	English
Singapore	English
Slovak Republic	Slovak or Hungarian
Slovenia	Slovene
Sweden	Swedish
Turkey	Turkish
Total Number of Countries Included	33

[1] Although 35 countries participated in the 2001 PIRLS test, only 33 are included in the sample used in this paper, since parent survey responses are unavailable for Morocco and the United States.

[2] Data source: *PIRLS 2001 User Guide for the International Database, Supplement Two, Documentation of National Adaptations of Background Questionnaire Items.*

Table 3.8: Summary Statistics on Variables Used

Variable	Observations	Mean	Standard Deviation
Raw Score on All Items ($score_{ist}$)	19950	51.06	9.61
Student Characteristics			
Boy	19950	0.49	0.50
Speak Language of Test when Little	19950	0.88	0.32
Speak Language of Test at Home	19950	0.94	0.24
Hours Spent Watching TV per Day			
More than 5 Hours	19950	0.10	0.30
Between 3 to 5 Hours	19950	0.11	0.32
Between 1 to 3 Hours	19950	0.32	0.47
Up to an Hour	19950	0.34	0.47
None	19950	0.07	0.27
Immigrant	19950	0.14	0.35
Family Background			
Father being Immigrant	19950	0.85	0.36
Mother being Immigrant	19950	0.84	0.36
Parent's Reading Habit			
More than 10 Hours per Week	19950	0.16	0.36
Between 6 to 10 Hours per Week	19950	0.22	0.41
Between 1 to 5 Hours per Week	19950	0.35	0.48
Less than an Hour per Week	19950	0.13	0.34
Annual Family Income			
More than $60,000	19950	0.10	0.30
$50,000 to $59,999	19950	0.05	0.22
$40,000 to $49,999	19950	0.07	0.26
$30,000 to $39,999	19950	0.08	0.27
$20,000 to $29,999	19950	0.09	0.29
Below $20,000	19950	0.14	0.34

Table 3.8: Summary Statistics on Variables Used, continued

Variable	Observations	Mean	Standard Deviation
Reading-Related Activities			
Read Book	19950	0.79	0.41
Tell Story	19950	0.79	0.41
Sing Song	19950	0.74	0.44
Play Alphabet Toy	19950	0.65	0.48
Computer	19950	0.19	0.39
Play Word Game	19950	0.64	0.48
Write Letters or Words	19950	0.76	0.43
Read Signs and Labels Aloud	19950	0.69	0.46
Watch TV Reading Programs	19950	0.53	0.50
Teacher Characteristics			
Experience	18624	18.56	16.03
Male	19305	0.17	0.38
Certificate	19305	0.84	0.37
School Characteristics			
Students from Economically Disadvantaged Background			
More than 50%	19950	0.17	0.37
26 to 50%	19950	0.18	0.38
11 to 25%	19950	0.25	0.43
0 to 10%	19950	0.37	0.48

Table 3.8: Summary Statistics on Variables Used, continued

Variable	Observations	Mean	Standard Deviation
Before First Grade:			
Recognize Most of Alphabet			
More than 75%	19950	0.22	0.42
51 to 75%	19950	0.16	0.36
25 to 50%	19950	0.18	0.38
Less than 25%	19950	0.37	0.48
Read Some Words			
More than 75%	19950	0.16	0.37
51 to 75%	19950	0.13	0.34
25 to 50%	19950	0.19	0.39
Less than 25%	19950	0.48	0.50
Read Sentences			
More than 75%	19950	0.09	0.29
51 to 75%	19950	0.08	0.27
25 to 50%	19950	0.12	0.33
Less than 25%	19950	0.66	0.47
Write Alphabet			
More than 75%	19950	0.18	0.39
51 to 75%	19950	0.15	0.35
25 to 50%	19950	0.21	0.41
Less than 25%	19950	0.39	0.49
Write Some Words			
More than 75%	19950	0.14	0.35
51 to 75%	19950	0.11	0.31
25 to 50%	19950	0.17	0.38
Less than 25%	19950	0.55	0.50

[1] All variables, except for *Raw Score on All Items* ($score_{ist}$) and *Experience*, are dummy variables, containing a value of either 1 or 0.

[2] Author's computation using data and answers given in the 2001 PIRLS student, parent (home), teacher, and principal (school) questionnaires. These summary statistics reflect individuals who have no missing values reported in the variables of interest.

3.6 Appendix: Contents Covered in the PIRLS Test

The goal of administering the Progress in International Reading Literacy Study (PIRLS) test is to assess the general reading ability among nine year-olds at a global level. In particular, the test aims at evaluating a student's proficiency in understanding literary meanings, as well as drawing inferences from passages. As a result, two broad types of questions are covered: literary and informational. The literary section refers to testing whether a student can comprehend ideas being covered in the passage, while the informational section asks him or her to use information given in the text to draw inferences.

The PIRLS test pool consists of 8 reading blocks, 4 being literary and 4 being informational. Each block contains one or several passages, followed by questions that test a student's understanding of the passage(s). In each test booklet, only 2 reading blocks (one from each of the broad items, literary and informational) are covered, since completing all 8 blocks would be too exhaustive for the test-taker. As a result, there are 10 different booklets assembled, each containing a combination of 2 out of the 8 reading blocks. Sample reading passages can be found in Appendix C in Mullis, et al. (2003), *PIRLS 2001 International Report: IEA's Study of Reading Literacy Achievement in Primary Schools in 35 Countries*, or at *http://pirls.bc.edu/pirls2001/pdf/P1_IR_C.pdf*.

Bibliography

[1] *1997 NAICS United States Structure, Including Relationship to 1987 U.S. SIC* (2000), Washington, D.C.: Census Bureau.

[2] Adams, R.J. (2003). "Response to 'Cautions on OECD's Recent Educational Survey (PISA),' " *Oxford Review of Education*, 29(3): 377-389.

[3] Altonji, J.G. and D. Card (1989). "The Effects of Immigration on the Labor Market Outcomes of Natives," NBER Working Paper No. 3123.

[4] Ammermüller, A. and P. Dolton (2006). "Pupil-Teacher Gender Interaction Effects on Scholastic Outcomes in England and the USA," ZEW Discussion Paper No. 06-060.

[5] Ammermüller A. and J. Pischke (2006). "Peer Effects in European Primary Schools: Evidence from PIRLS," IZA Discussion Paper No. 2077.

[6] Bacolod, M., B.S. Blum, and W.C. Strange (2007). "Skills in the City," University of Toronto Working Paper.

[7] Barro, R.J. and J.W. Lee (1996). "International Measures of Schooling Years and Schooling Quality," *The American Economic Review*, 86(2): 218-223.

[8] Berman, E., K. Lang, and E. Siniver (1999). "Language-Skill Complementarity: Returns to Immigrant Language Acquisition," Boston University Working Paper.

[9] Betts, J.R. (1995). "Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth," *The Review of Economics and Statistics*, 77(2): 231-250.

[10] Betts, J.R. and M. Lofstrom (2000). "The Educational Attainment of Immigrants: Trends and Implications," in G.J. Borjas, ed., **Issues in the Economics of Immigration**. Chicago: University of Chicago Press for National Bureau of Economic Research: 51-115.

[11] Bishop, J. (1992). "The Impact of Academic Competencies of Wages, Unemployment, and Job Performance," *Carnegie-Rochester Conference Series on Public Policy*, 37: 127-194.

[12] Bleakley, H. and A. Chin (2003). "Language Skills and Earnings: Evidence from Childhood Immigrants," CCIS Working Paper No. 87.

[13] Borjas, G.J. (1985). "Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants," *Journal of Labor Economics*, 3(4): 463-489.

[14] Borjas, G.J. (1987). "Self-Selection and the Earnings of Immigrants," *The American Economic Review*, 77(4): 531-553.

[15] Borjas G.J. (1991). "Ethnic Capital and Intergenerational Mobility," NBER Working Paper No. W3788.

[16] Borjas, G.J. (1995). "Assimilation and Changes in Cohort Quality Revisited: What Happened to Immigrant Earnings in the 1980s?" *Journal of Labor Economics*, 13(2): 201-245.

[17] Borjas, G.J. (1999). "The Economic Analysis of Immigration," in O. Ashenfelter and D. Card, eds., **Handbook of Labor Ecoonomics**. Amsterdam: North-Holland: 1697-1760.

[18] Borjas, G.J. (2001). **Heaven's Door: Immigration Policy and the American Economy**. New Jersey: Princeton University Press.

[19] Bratsberg, B. and D. Terrell (2002). "School Quality and Returns to Education of U.S. Immigrants," *Economic Inquiry*, 40(2): 177-198.

[20] Brewer, D.J., E.R. Eide, and R.G. Ehrenberg (1999). "Does It Pay to Attend an Elite Private College? Cross-Cohort Evidence on the Effects of College Type on Earnings," *Journal of Human Resources*, 34(1), 104-123.

[21] Cavanagh, S. (2008). "Playing Games in Classroom Helping Pupils Grasp Math." *Education Week*, April 30: 10.

[22] Chiquiar, D. and G.H. Hanson (2005). "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States," *Journal of Political Economy*, 113(2): 239-281.

[23] Chiswick, B.R. (1978). "The Effect of Americanization on the Earnings of Foreign-Born Men," *Journal of Political Economy*, 86(5): 897-921.

[24] Chiswick, B.R. (1980). "An Analysis of the Economic Progress and Impact of Immigrants," **Final Report to the U.S. Department of Labor**. Washington: U.S. Department of Labor.

[25] Chiswick, B.R. (1999). "Are Immigrants Favorably Self-Selected?" *AEA Papers and Proceedings*, 89(2): 181-185.

[26] Chiswick, B.R. and P.W. Miller (1995). "The Endogeneity Between Language and Earnings: International Analyses," *Journal of Labor Economics*, 13(2): 245-287.

[27] Chiswick, B.R. and P.W. Miller (1998). "Language Skills and Earnings Among Legalized Aliens," *Journal of Population Economics*, 12(1): 63-89.

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

[29] Church, J. and I. King (1993). "Bilingualism and Network Externalities," *Canadian Journal of Economics*, 26(2): 337-345.

[30] Currie, J. and D. Thomas (1995). "Does Head Start Make a Difference?" *The American Economic Review*, 85(3): 341-364.

[31] Deininger, K. and L. Squire (1996). "A New Data Set Measuring Income Inequality," *World Bank Economic Review*, 10(3): 565-591.

[32] Deininger, K. and L. Squire (1996). *Measuring Income Inequality Database*. Herndon: World Bank.

[33] **Digest of Educational Statistics 2002** (2002). Washington: National Center for Education Statistics.

[34] Feenstra, R.C. (1996). "U.S. Imports, 1972-1994: Data and Concordances," NBER Working Paper No. 5515. Related data available on *The Center for International Data at UC Davis* website, *http://cid.econ.ucdavis.edu/*.

[35] Feenstra, R.C., J. Romalis, and P.K. Schott (2002). "U.S. Imports, Exports and Tariff Data, 1989-2001," NBER Working Paper No. 9387. Related data available on *The Center for International Data at UC Davis* website, *http://cid.econ.ucdavis.edu/*.

[36] Friedberg, R.M. (2000). "You Can't Take It with You? Immigrant Assimilation and the Portability of Human Capital," *Journal of Labor Economics*, 18(2): 221-251.

[37] Gonzales, P., C. Calsyn, L. Jocelyn, K. Mak, D. Kastberg, S. Arafeh, T. Williams, and W. Tsen (2000). **Pursuing Excellence: Comparisons of International Eighth-Grade Mathematics and Science Achievement from a U.S. Perspective, 1995 and 1999**. Jessup: Education Publications Center.

[38] Gonzalez, E.J. and A.M. Kennedy (eds.) (2003). *PIRLS 2001 User Guide for the International Database*. Chestnut Hill, MA: International Study Center, Boston College.

[39] Gould, D.M. (1994). "Immigrant Links to the Home Country: Empirical Implications for U.S. Bilateral Trade Flows," *The Review of Economics and Statistics*, 76(2): 302-316.

[40] Hanushek, E.A. (1998). "The Evidence on Class Size," University of Rochester Occasional Paper No. 98-1.

[41] Hanushek, E.A. (2003). "The Failure of Input-based Schooling Policies," *The Economic Journal*, 113(February): F64-F98.

[42] Hanushek, E.A. and D.D. Kimko (2000). "Schooling, Labor-Force Quality, and the Growth of Nations," *The American Economic Review*, 90(5): 1184-1208.

[43] Hanushek, E.A. and J.A. Luque (2003). "Efficiency and Equity in Schools Around the World," *Economics of Education Review*, 20(5): 481-502.

[44] Hanushek, E.A. and L. Woessmann (2005). "Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence Across Countries," CESifo Working Paper No. 1415.

[45] Heckman, J., A. Layne-Farrar, and P. Todd (1996). "Human Capital Pricing Equations with an Application to Estimating the Effect of Schooling Quality on Earnings," *The Review of Economics and Statistics*, 78(4): 562-610.

[46] Hoxby, C.M. (2000). "The Effects of Class Size on Student Achievement: New Evidence from Population Variation," *Quarterly Journal of Economics*, 115(4): 1239-1285.

[47] Hoxby, C.M. (2001). "The Return to Attending a More Selective College: 1960 to the Present," in M. Devlin and J. Meyerson, eds., **Forum Futures: Exploring the Future of Higher Education, 2000 Papers**. New York: Jossey-Bass Inc.: 13-42.

[48] Hungerford, T. and G. Solon (1987). "Sheepskin Effects in the Returns to Education," *The Review of Economics and Statistics*, 69(1): 175-177.

[49] Hutchinson, W.K. (2002). "Does Ease of Communication Increase Trade? Commonality of Language and Bilateral Trade," *Scottish Journal of Political Economy*, 49(5): 544-556.

[50] "International Electricity Imports, All Countries, 1980-2005 for the *International Energy Annual 2005*" (2005). Washington, D.C.: Energy Information Administration. *http://www.eia.doe.gov/emeu/international/electricitytrade.html*

[51] Jaeger, D.A. and M.E. Page (1996). "Degrees Matter: New Evidence on Sheepskin Effects in the Returns to Education," *The Review of Economics and Statistics*, 78(4): 733-740.

[52] Jepsen, C. and S. Rivkin (2002). *Class Size Reduction, Teacher Quality, and Academic Achievement in California Public Elementary Schools*. San Francisco: Public Policy Institute of California.

[53] Jürges H. and K. Schneider (2007). "Fair Ranking of Teachers," *Empirical Economics*, 32: 411-431.

[54] Kahn, S. and K. Lang (1996). "Hours Constraints and the Wage/Hours Locus," *Canadian Journal of Economics/Revue Canadienne d'Economique*, 29(Part 1): S71-S75.

[55] Krueger, A.B. and L.H. Summers (1988). "Efficiency Wages and the Inter-Industry Wage Structure," *Econometrica*, 56(2): 259-293.

[56] Lang, K. (1986). "A Language Theory of Discrimination," *Quarterly Journal of Economics*, 101(2): 363-382.

[57] Lazear, E.P. (1998). "Diversity and Immigration," NBER Working Paper No. 6535.

[58] Lazear, E.P. (1999). "Culture and Language," *Journal of Political Economy*, 107(6): S95-S126.

[59] Leon, A. (2003). "Does 'Ethnic Capital' Matter? Identifying Peer Effects in the Intergenerational Transmission of Ethnic Differentials," Dissertation Chapter, MIT.

[60] Lucas, R.E.B. (1977). "Hedonic Wage Equations and Psychic Wages in the Returns to Schooling," *The American Economic Review*, 67(4): 549-558.

[61] Martin, M.O., I.V.S. Mullis, A. M. and Kennedy (2003). *PIRLS 2001 Technical Report*. Chestnut Hill, MA: International Study Center, Boston College.

[62] Meghir, C. and M. Palme (2004). "Educational Reform, Ability and Family Background," IFS Working Paper No. WP04/10.

[63] Mincer, J. (1970). "The Distribution of Labor Incomes: A Survey with Special Reference to the Human Capital Approach," *Journal of Economic Literature*, 8(1): 1-26.

[64] Mincer, J. (1974). **Schooling Experience and Earnings**. New York: NBER.

[65] Mullis, I.V.S., M.O. Martin, E.J. Gonzalez, K.D. Gregory, R.A. Garden, K.M. O'Connor, S. J. Chrostowski, and T.A. Smith (2000). ***TIMSS 1999 International Mathematics Report****: Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*. Chestnut Hill, MA: International Study Center, Boston College.

[66] Mullis, I.V.S., M.O. Martin, E. J. Gonzalez, and A. M. Kennedy (2003). *PIRLS 2001 International Report: IEA's Study of Reading Literacy Achievement in Primary Schools in 35 Countries*. Chestnut Hill, MA: International Study Center, Boston College.

[67] Mullis, I.V.S., M.O. Martin, A. M. Kennedy, and C.L. Flaherty (eds.) (2002). *PIRLS 2001 Encyclopedia: A Reference Guide to Reading Education in the Countries Participating in IEA's Progress in International Reading Literacy Study (PIRLS)*. Chestnut Hill, MA: International Study Center, Boston College.

[68] Ogle, L., A. Sen, E. Pahlke, L. Jocelyn, D. Kastberg, S. Roey, and T. Williams (2003). *International Comparisons in Fourth-Grade Reading Literacy: Findings from the Progress in International Reading Literacy Study (PIRLS) of 2001* (NCES 2003-073). U.S. Department of Education, NCES. Washington, DC: U.S. Government Printing Office.

[69] *Percent of Labor Force Age 25-34 By Education* (2002). Washington, D.C.: Bureau of Labor Statistics.

[70] *PIRLS 2001 User Guide for the International Database Supplement One: International Version of the Background Questionnaires* (2003). Chestnut Hill, MA: International Study Center, Boston College. *http://timss.bc.edu/pirls2001i/PIRLS2001_Pubs_UG.html*

[71] *PIRLS 2001 User Guide for the International Database Supplement Two: Documentation of National Adaptations of Background Questionnaire Items* (2003). Chestnut Hill, MA: International Study Center, Boston College. *http://timss.bc.edu/pirls2001i/PIRLS2001_Pubs_UG.html*

[72] Psacharopoulos, G. (1993). "Returns to Investment in Education, A Global Update," Policy Research Working Papers No. 1067.

[73] Psacharopoulos, G. (1995). "The Profitability of Investment in Education: Concepts and Methods," HCO Working Papers No. 63.

[74] Rauch, J.E. and V. Trindade (1999). "Ethnic Chinese Networks in International Trade," NBER Working Paper No. 7189.

[75] Rose, H. and J.R. Betts (2001). *Math Matters: The Link Between High School Curriculum, College Graduation, and Earnings*. San Francisco: Public Policy Institute of California.

[76] Ruggles, S., M. Sobek et al. (2003). *Integrated Public Use Microdata Series: Version 3.0*. Minneapolis: Historical Census Projects, University of Minnesota.

[77] Schnepf, S.V. (2007). "Immigrants' Educational Disadvantage: An Examination Across Ten Countries and Three Surveys," *Journal of Population Economics*, 20: 527-545.

[78] Tiebout, C. (1956). "A Pure Theory of Local Expenditures," *Journal of Political Economy*, 64(5): 416-424.

[79] Uusitalo, R. (1999). "Return to Education in Finland," *Labour Economics*, 6: 569-580.

[80] Woessmann, L. (2003). "Educational Production in East Asia: The Impact of Family Background and Schooling Policies on Student Performance," Kiel Working Paper No. 1152.

[81] Woessmann, L. and T. Fuchs (2005). "Families, Schools, and Primary-School Learning: Evidence for Argentina and Colombia in an International Perspective," World Bank Policy Research Working Paper No. 3537.

[82] **The World Factbook 1994** (1994). Washington, D.C.: Central Intelligence Agency.

[83] **The World Factbook 2007** (2007). Washington, D.C.: Central Intelligence Agency.

[84] *http://www.indo.com* provides measure on the distance between various countries' capitals and Washington, D.C.