DISCUSSION PAPER SERIES

IZA DP No. 7880

International Migration and the Economics of Language

Barry R. Chiswick
Paul W. Miller

January 2014

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

International Migration and the Economics of Language

Barry R. Chiswick
George Washington University
and IZA

Paul W. Miller
Curtin University
and IZA

Discussion Paper No. 7880
January 2014

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 7880
January 2014

ABSTRACT

International Migration and the Economics of Language*

This paper provides a review of the research on the 'economics of language' as applied to international migration. Its primary focuses are on: (1) the effect of the language skills of an individual on the choice of destination among international (and internal) migrants, both in terms of the ease of obtaining proficiency in the destination language and access to linguistic enclaves, (2) the determinants of destination language proficiency among international migrants, based on a model (the three E's) of Exposure to the destination language in the origin and destination, Efficiency in the acquisition of destination language skills, and Economic incentives for acquiring this proficiency, (3) the consequences for immigrants of acquiring destination language proficiency, with an emphasis on labor market outcomes, and in particular earnings. Factors that are considered include age, education, gender, family structure, costs of migration, linguistic distance, duration in the destination, return migration, and ethnic enclaves, among others. Analyses are reported for the immigrant experiences in the US, Canada, Australia, the UK, Germany, Israel and Spain.

JEL Classification: J15, J24, J31, J61

Keywords: immigrants, language, bilingualism, human capital, earnings

Corresponding author:

Barry R. Chiswick
Department of Economics
George Washington University
2115 G Street, NW
Monroe Hall 340
Washington, DC 20052
USA
E-mail: brchis@uic.edu

* Chapter prepared for Barry R. Chiswick and Paul W. Miller, eds. *Handbook on the Economics of International Migration*, Elsevier Inc (forthcoming).
We appreciate the research assistance of Marina Gindelsky and Derby Voon.

International Migration and the Economics of Language

1. Introduction

This chapter provides an overview of research on the economics of language, as applied to international migration.[1] The 'Economics of Language' is the study of the determinants and consequences of language proficiency using the methodology and tools of economics.

The beginning of interest by economists in language is usually attributed to Jacob Marschak (1965), who was concerned with the efficiency of communication. Using an evolutionary approach, those aspects of language that were beneficial in facilitating communication would survive, and those aspects (and languages) which were not efficient or effective would tend to disappear. In this framework languages tend to evolve over time. And just as mutations lead to the evolution of different species of plants and animals, isolation and language drift tend to promote the development of new dialects and languages. This approach did not generate much of a literature in economics on the evolution of languages.

It did, however, stimulate interest in language usage in bilingual and multilingual countries (e.g., Canada, Switzerland). Which language or languages become dominant, especially for economic activity, and who learns which language? In some countries, regional or indigenous minority group languages have disappeared or appear to be on the verge of vanishing (e.g., consider Celtic in Ireland and Scotland, the Sami language in Sweden, and indigenous peoples' languages in countries of overseas settlement such as the US, Canada, Australia, and Brazil). While in others, attempts have been made in recent decades with various degrees of

[1] This chapter is a development of Barry R. Chiswick, 'The Economics of Language Learning for Immigrants: An Introduction and Overview' in T.G. Wiley, J.S. Lee and R.W. Rumberger (eds.), *The Education of Language Minority Immigrants in the United States*, Bristol, UK: Multilingual Matters, 2009, pp. 72-91.

success to reinvigorate traditional languages (e.g., Catalonian in Spain and Welsh in Wales).

Following several decades of low rates of international migration due to two world wars, the Great Depression, and severe immigration restrictions in many major destinations, international migration started increasing in the 1950s, and it has continued to increase in each decade since. This migration led to an influx of people in various destinations who were not familiar with the primary or dominant language of the destination. This led to interest by economists in the nexus between language and immigration. In addition to advances in economic theory and the development of testable models, economists now had better (not perfect, but better) micro data to estimate their models and test their hypotheses regarding language and international migration.

Most of the research in the economics of language focuses on what can be described as microeconomics, that is, the behavior of individuals. The approach taken has been to view language skills as a form of 'human capital'. The concept of human capital became important in the 1960s, with the emphasis on schooling, on-the-job training, health and information, all of which transform the person, and migration, which transforms the person's location (Schultz, 1962). It was only since the 1980s, however, that economists have viewed immigrant language skills as a form of human capital and analyzed it in this context (Carliner, 1981; McManus et al., 1983; Tainer, 1988). This interest arose as a result of the rapid growth of the non-English speaking portion of the increasing immigrant flows into the US and Canada, the emerging interest among economists in the determinants of the adjustment of immigrants to the host society, and the growing interest in the application of human capital theory (Becker, 1964).

Language skills satisfy the three requirements for human capital in that they are productive, costly to produce, and embodied in the person. First, a person's proficiency in the language of the area in which he or she lives is productive in the labor market. Those who speak/read the local language will find it easier to obtain a job and will generally be more productive on the job. In addition, language skills are productive in consumption activities. Those proficient in the local language will be more efficient in the search for higher quality goods and services and at lower prices. Any monolingual English speaker in the Chinese countryside quickly learns this proposition. Immigrants who do not speak the language of the broader society also find that their social and information networks are confined to their immigrant/linguistic enclave, rather than having a wider range. These benefits from proficiency provide economic and social incentives for immigrants to learn the host country's language.

Second, acquiring language proficiency is not without costs. Immigrants spend a considerable amount of their own time and money (for language training schools, books, etc.) to become proficient in their new country's language. Acquiring language skills is not costless even for infants. Even if their own time has no economic value, the time of their parents or other caregivers in speaking and reading to the child is not costless. The costs involved in an immigrant learning a new language would be influenced by several factors, including the person's value of time (wage rate), the person's age, exposure to the destination language (as distinct from being able to avoid its use by living and working in a linguistic enclave) and the 'distance' between the person's mother tongue and the language of the destination, among other factors.

Finally, language skills are embodied in the person. Unlike owning physical capital (such

as a truck), but like learning to play a piano, language skills cannot be separated from the person.

The idea that language skills are both productive and costly to acquire is not new, but rather at least thousands of years old. See Box 1, which relates the story of the Tower of Babel from the Biblical book of Genesis (Chapter 11, verses 1-9). The Tower of Babel provided a biblical explanation for the diversity of languages and the scattering of people: 'If, as one people with one language ... then nothing that they may propose to do will be out of their reach'. When their speech was 'confounded' and they were scattered, they could no longer cooperate and they became less productive (Tanakh, 1985).[2]

Box 1: Tower of Babel

Everyone on earth had the same language and the same words. And as they migrated from the east, they came upon a valley in the land of Shinar and settled there.

They said to one another, 'Come, let us make bricks and burn them hard'.— Brick served them as stone, and bitumen served them as mortar. — And they said, 'Come let us build a city, and a tower with its top in the sky, to make a name for ourselves; else we shall be scattered all over the world'.

The LORD came down to look at the city and tower that man had built, and the LORD said, 'If, as one people with one language for all, this is how they have begun to act, then nothing that they may propose to do will be out of their reach. Let us, then, go down and confound their speech there, so that they shall not understand one another's speech.' Thus the LORD scattered them from there over the face of the whole earth; and they stopped building the city. This is why it was called Babel, because there the LORD confounded the speech of the whole earth; and from there the LORD scattered them over the face of the whole earth. (Genesis, 11, 1-9)

Source: *Tanakh: The Holy Scriptures*, pp.16-17 (1986) Philadelphia, Jewish Publication Society.

[2]An important issue currently facing the United States, and most of the highly developed economies, is the inverse of the Tower of Babel story. Immigration is resulting in the coming together of diverse peoples originally speaking a variety of languages who then merge over time into a common culture and a common language, even if they may also retain the languages of their origins.

The structure of this chapter is as follows. In the next section the three major aspects of the relation between language and international migration are outlined, and a broad overview of the methodology adopted in the empirical research is provided. This is followed by separate sections on each of the three major research themes, where the conceptual frameworks followed in the literature are presented, and empirical evidence discussed. The chapter ends with a summary of the findings and a discussion of gaps in the literature which warrant further research.

2. Research Issues and Methodology

Much of the research on the economics of language as applied to international migration has focused on three main issues. The first matter addressed is the links between language background and the decision to migrate and the choice of destination by migrants. The language factors that are important include knowledge of a language that is used in the destination country, as well as knowledge of a language which, because it is linguistically close to a language used in the destination, makes learning the destination country language easier. The second issue concerns the determinants of proficiency in the primary or dominant language of the country of destination, including in the labor market; although the model and the methodology can be, and have been, applied to non-migrants who are linguistic minorities and native-born bilingual speakers. The third major concern covers the consequences for immigrants of obtaining proficiency in the dominant language. The consequences of language proficiency that has received the most attention have been in the labor market, particularly earnings. There are, however, other consequences which have not received much attention from economists or other social scientists. The primary focus on earnings has arisen in part because of interest in economic

well-being and in part because of the greater availability of data on earnings.

Knowing the dominant language makes a person more efficient in the consumption of goods and services (higher quality and lower prices for goods and services). Investments in other forms of human capital, such as schooling and job training, are likely to be more productive if one can communicate in the dominant language in school and in the labor market. Knowing the dominant language of the destination can also increase the efficiency of parenting. Parents who are proficient in the dominant language can be more effective in teaching the language and culture of the destination to their children, which would be a benefit to them in school and later in the job market. Language skills also have social benefits as they can expand the range of friendship networks beyond one's ethnic/linguistic enclave. Finally, civic involvement is enhanced with knowledge of the host country's language. Indeed, for the United States and many other countries, at least a basic knowledge of the destination language is required for immigrants to become citizens and acquire full political and economic rights. This brings about increased political empowerment. There is no doubt that these non-labor market consequences of dominant language proficiency are important. However, to have a manageable review we focus on the labor market consequences.

The particular models advanced by economists in relation to each of these research questions are discussed in Sections 3, 4 and 5. The testing of the models, or the estimation of the equations, relies on multivariate statistical (econometric) techniques.

In general, many economists believe in the importance of testing for the robustness of findings. One set of estimates from one data set may be insightful, but cannot determine whether the results are unique to that data, group, country or time period, or whether they are

generalizable across these dimensions. A hypothesis or model that is not robust, but is valid for only a unique group, time and place, is clearly of very limited value. On the other hand, one has much greater confidence in a hypothesis or model that is robust, that is, supported by analyses of diverse data sets.

The analyses reported below represent a synthesis of the findings on immigrants for different types of data, censuses and surveys, both cross sectional and longitudinal. They are for immigrants who have legal status, as well as those with an illegal or unauthorized status. Although the data analyses reported below are for the late 20th and early 21st centuries, they are for different data sets across several countries, primarily the United States, Australia, Canada, Germany and Israel, where the destination language is English in the first two, English and French in Canada, German in Germany, and Hebrew in Israel. The particular value of research on Israel and Germany in this context is that whereas English, and to a lesser extent French, is an international language of culture, business and science, which is often learned in school in the country of origin as a second language, this is less so for Hebrew for immigrants to Israel, or for German among the groups of immigrants covered in the empirical research for Germany. That the findings for Israel and Germany parallel those of the other countries is a test of the robustness of the model across destination languages (Chiswick, 1998; Chiswick and Repetto, 2001; Dustmann, 1994). In the case of the links between language background and destination language choice, many destination countries, most countries of origin in the world, and a wide range of languages, are covered.

There are several dimensions of language skills—oral (speaking and hearing) and literacy (reading and writing). Survey and census data on the language skills of immigrants almost

always rely on self-reported responses or responses provided by an adult household member. Although some data sets report responses to questions for immigrants on reading and writing, most of the data are regarding speaking skills, focusing on either the self-reported level of competency or identifying the languages spoken on a regular basis. Analyses using literacy skills show the same patterns as those using speaking skills, in part because the two are so highly correlated (Chiswick, 1991; Chiswick and Repetto 2001; Dustmann, 1994). The discussion here will be expressed in terms of speaking proficiency, unless noted otherwise.

3. Choice of Destination

Migration, whether internal to a country or international, is an investment in human capital. Unlike other forms of human capital, such as schooling, health and information which transform the person, migration transforms where the person lives or works. It is a form of human capital because migration is costly, is beneficial in either consumption or production (e.g., labor market work), and the migration per se cannot be separated from the person.

The costs of migration include out-of-pocket costs (sometimes referred to as direct costs) and foregone earnings (sometimes referred to as opportunity costs). The costs are far greater than merely the costs of moving oneself, family members, and household goods from one point (the origin) to another (the destination). The costs incurred in the origin include the separation from family, friends and a familiar environment. The set of skills acquired in the origin, including language skills, and which are useful in consumption and production activities, may not be equally useful in the destination. The transferability from the origin to the destination may be limited by geographic differences in technology, by custom, by occupational licensing, etc. The

transferability may also be limited by language differences, where language can be considered a technology for communication. A Chinese speaker in a monolingual English country would find that being a consumer who looks for higher quality goods and services at lower prices is more difficult than it was in China, or that the job search process is more difficult. Wages would be lower and employment conditions less desirable if linguistic disadvantages lower the workers' productivity on the job or increase workplace costs (e.g., accidents). Relevant language skills may be important for communicating with supervisors, peers, and subordinates, as well as suppliers and customers or clients.

As a result, part of the cost of migration is the lower earnings during the period of adjusting one's language skills, as well as the cost of direct investment in improving destination language skills (e.g., the opportunity cost and out-of-pocket cost of a language training program). The latter cost may be incurred prior to the migration, in anticipation of the move, or after migration.

The costs of language adjustment depend on many factors, which will be discussed in greater detail below. These costs will include the importance and the ease or difficulty of learning the language of the destination. Potential migrants need not consider just one destination, but may consider the range of potential or available destinations. Therefore, part of the calculation as to which destination is most preferred is the cost of destination language acquisition. The language acquisition costs are lower the 'closer' is the language of the origin to that of the destination. Thus, for an Australian, the linguistic cost of adjusting to a destination is cheaper if it is the UK rather than France, or for a Spaniard if the destination is Argentina, rather than Canada. The implication is that language differences are among the factors that influence

the choice of destination.

Much of the research on the links between language background and the choice of destination has used aggregate-level data. Two decisions have been studied: the choice of initial destination country; and internal migration in the years following arrival in the host country. Evidence from both streams of literature is reviewed.

3.1 *Choice of initial destination country*

The research on the links between language background and choice of the initial destination can be illustrated through detailed coverage of the study by Clark, Hatton and Williamson (2007), and then covering other studies more briefly.

Clark et al. (2007) try to account for the determinants of migration rates to the US by place of birth for 81 source countries from 1971 to 1998. They based their analysis around the following model:[3]

$$mig_j / pop_j = f(y_j / y_{US}, syr_j / syr_{US}, inq_j / inq_{US}, age_j, pov_j, dist_j, land_j, eng_j, stock_j / pop_j, policy\ variables)$$

where mig_j / pop_j is the flow of migrants from source country *j* to the US in a particular year, normalized by the population of the source country, y_j / y_{US} is the average (purchasing power parity adjusted) income in the source country *j* relative to that in the US, the terms in syr and inq capture, respectively, differences in average years of schooling and inequality in source country *j* and the US, *age* is the share of the population in the sending country aged 15-29, *pov* is

[3] Time subscripts are suppressed to simplify the notation.

the poverty rate in the origin country, *dist* is the geographic distance of the source country from the US, *land* denotes cases where the origin country is land-locked, *eng* is for where the source country is predominately English speaking, and the *stock* variable captures the number of previous immigrants from the source country. The Clark et al. (2007) model also contained policy variables, for the number of visas available in the different visa classes, and institutional factors, such as the US 1986 Immigration Reform and Control Act (IRCA) legalization program. Hence, it is seen that the model of equation (1) captures economic and demographic drivers of migration in the first five terms, with the remaining terms representing costs and policy parameters. In particular, it was argued that the cost of migration to the US would be higher if the sending country was not predominantly English speaking.

Clark et al.'s (2007) results show that the English-speaking sending countries were associated with a statistically significant higher migration rate to the US in the models that did not include the immigrant stock variables. In other words, having an English language background matters when it comes to understanding rates of migration to an English-speaking country. This effect, however, was not statistically significant where the stock of previous migrants from the sending country was included in the estimating equation. As noted by Clark et al. (2007. p.267) 'Since the immigrant stock reflects past immigration, it captures much of the effect of slow moving fundamentals over the longer term'. Included in these fundamentals would be the country-specific cost factors that the English-speaking background variable is used to represent.

Approaches similar to the model that Clark et al. (2007) applied to English-speaking countries have been adopted by other researchers when considering migration flows into

countries characterized by a number of official languages. For example, Karemera, Oguledo and Davis (2000) examine migration flows to the US, using a common language (English) variable along the lines of Clark et al. (2007), and to Canada, where they use a variable for the language of the origin country being either English or French. Karemera et al. (2000) report, however, that language commonality was not a significant determinant of migration rates to either the US or Canada over the decade 1976-1986. One of the main differences between the models of Karemera et al. (2000) and Clark et al. (2007) is that the former includes a set of dichotomous variables for region of origin.[4]

The coverage of common languages has been extended further by Pedersen, Pytlikova and Smith (2008). They examined gross migration flows from 129 countries into 22 OECD destination countries over 1990-2000. The 22 OECD countries have a number of official languages (English, German, Spanish, Italian, etc.), and so the English language variable in the model of Clark et al. (2007) was replaced by a common language variable in the Pedersen et al. (2008) study. This variable was set equal to one where there is a common language between the origin and destination, and it is set equal to zero where there is no common language, with the *Ethnologue: Languages of the World* (2009) being used to classify pairs of countries (see Box 2). An additional feature of this study is that it also includes a dummy variable for countries that were ever in a colonial relationship. Presumably this variable is correlated with the common language variable, though this was not examined in the study. Nevertheless, Pedersen et al. (2008) report that the common language variable was an important influence on migration flows,

[4] The studies in this field are characterized by differences in the choice of dependent variable (whether the immigrant flow is normalized by the source country population, and whether flows or stocks are used), the selection of independent variables, as well as the lag structure of variables. We do not discuss these specification issues in detail here.

being significant in five of the eight models presented in their main set of results.

Box 2: Ethnologue: Languages of the World (16th edition of 2009)

This is a comprehensive reference volume that catalogues 7,413 languages, including details on the 6,909 known living languages in the world at the time of writing. Each language is part of a language family (that is, its linguistic lineage is provided). For example, the linguistic lineage for English is, from largest grouping to smallest, Indo-European—Germanic—Germanic West—English. This type of connectivity between languages has been used in various ways by researchers to construct a measure of linguistic distance or proximity. It can be illustrated using the algorithm proposed by Adsera and Pytlikova (2012). Thus, they construct their variable as follows:

> First we defined weights: the first equal to 0.1 if two languages are related at the most aggregated linguistic tree level, e.g. Indo-European versus Uralic (Finnish, Estonian, Hungarian); the second equal to 0.15 if two languages belong to the same second-linguistic tree level, e.g. Germanic versus Slavic languages; the third equal to 0.20 if two languages belong to the same third-linguistic tree level, e.g. Germanic West vs. Germanic North languages; and the fourth equal to 0.25 if both languages belong to the same fourth level of a linguistic tree family, e.g. Scandinavian West (Icelandic) vs. Scandinavian East (Danish, Norwegian and Swedish), German vs. English, or Italo-Western (Italian, French, Spanish, Catalan and Portuguese) vs. Romance Eastern (Romanian). Then, we constructed the linguistic proximity index as a sum of those four weights, and we set the index equal to 0 if two languages did not belong to any common language family, and equal to 1 if the two countries had a common language. Thus the linguistic proximity index equals 0.1 if two languages are only related at the most aggregated linguistic tree level, e.g. Indo-European languages; 0.25 if two languages belong to the same first- and second-linguistic tree level, e.g. Germanic languages; 0.45 if two languages share the same first- up to third-linguistic tree level, e.g. Germanic North languages; and 0.7 if both languages share all four levels of a linguistic tree family, e.g. Scandinavian East (Danish, Norwegian and Swedish). (Adsera and Pytlikova, 2012, p.12).

Obviously, other scales can be derived using this information, see for example, Belot and Hatton (2012), and this has been argued by Isphording and Otten (2012) to be a weakness of the approach. They argue, for example 'This linguistic-tree approach has to deal with strong cardinality assumptions, and arbitrarily chosen parameters. Additionally, the approach offers only low variability between different language pairs and is difficult to implement for isolated languages such as Korean' (Isphording and Otten, 2012, p.5). Both Adsera and Pytlikova, (2012) and Belot and Hatton (2012) have used these data for measures of linguistic distance in immigration research. Other researchers, for example Pedersen et al. (2008), have used this source to compile a common language dummy variable.

Source: Lewis, M. P. (ed.), (2009). *Ethnologue: Languages of the World*, Sixteenth edition. Dallas, Tex.: SIL International. Online version: http://www.ethnologue.com/.

Mayda (2010) and Ortega and Giovanni (2009) also examine the determinants of immigration flows to OECD countries. Mayda (2010) covered 14 OECD countries, 79 sending countries, and the time period 1980-1995, whereas Ortega and Giovanni (2009) cover 14 OECD countries, 73 countries of origin, and the longer time period of 1980 to 2005. Neither study found language background to be of importance to the explanation of migration flows. Ortega and Giovanni (2009, p.14) argued that 'This is hardly surprising as most of the large migratory flows to the OECD (except for Mexico-US) take place between countries that do not share a land border or a common language'. Moreover, Mayda (2010, p.1263) notes that 'The impact of a common language, though of the right sign, is not statistically significant and, surprisingly, past colonial relationships do not appear to affect migration rates (this is true whether common language and colony are entered in the regression together or one at a time)'.

A more recent study that reports that a common language is important to understanding international migration is Grogger and Hanson (2011). They study the stock of immigrants in 15 high-income OECD countries in 2000, and employ both common language and an English-speaking destination country variables. Their model has separate equations for the scale of international migration flows of low-skilled (primary educated) and high-skilled (tertiary educated) workers, for the selection on the basis of the skills of immigrants (primary or tertiary educated from a particular origin country in the destination, compared to those who remained in the origin) and for sorting across destinations (the mix of immigrants across the destination countries). The preferred set of results for the scale equation shows that immigrants are more likely to be from a country that has a language in common with the destination country. The results for the selection equation indicate that immigrants that move to a country with a language

in common with the origin country are positively selected in terms of education levels, while the findings for the sorting equation suggest that destinations that have a language in common with the origin country attract more highly skilled immigrants. Similar findings are reported for the English-speaking destination country variable.

Beine et al. (2011) examined matters similar to Grogger and Hanson (2011), but from the perspective of changes in immigrant stocks between 1990 and 2000, using data on 195 source countries and 30 OECD countries. Their common language variable was a highly significant determinant of both low-skill and high-skill migration flows. A common language was also associated with a statistically significant positive effect on the skill ratio of these migration flows.

Hence, while the results from the research that has been based on the importance of a common language to understanding international immigration are somewhat mixed, the findings predominately suggest that a common language is an important determinant of the scale and mix of the migration flows.

There have been a number of developments in the analyses of the effect of language on the choice of destination. Of greatest relevance to this review are the studies that use a measure of 'linguistic distance' in place of the common language variable. This development reflects the fact that migration costs are lower where the migrant can easily learn the language of the destination country, for example, for migrants with a mother tongue that is linguistically close to the dominant language of the destination (see the next section).

Belot and Hatton (2012) examine the characteristics of migration for 70 source countries and 21 OECD destination countries for 2000/2001. While this study continues the theme of the

research by Grogger and Hanson (2011) and Beine et al. (2011) by focusing on educational selectivity, the key feature from our perspective is their linguistic proximity variable. They derive this measure from the language family information presented in the 'Ethnologue: Languages of the World' (See Box 2). As constructed by these authors, the linguistic proximity variable has values from 1 to 5, according to the number of common nodes in the linguistic tree between the closest official languages of pairs of countries.

The Ethnologue common node measure of Belot and Hatton (2012) was statistically significant and positive in their skill selection equation, a finding that the authors argue shows that the transferability of human capital might be easier when the linguistic gap is less, and hence immigrants can readily learn the dominant language of the destination.[5]

There might be a suspicion that the stronger results obtained with the seemingly superior measure of linguistic distance in Belot and Hatton (2012) is linked to their focus on stocks of immigrants. After all, Grogger and Hanson (2011) found that the conventional common language variable was highly significant in their model that has a focus on stocks, rather than annual flows. This matter can be addressed through reviewing the research of Belot and Ederveen (2012). They examine a panel of 22 OECD countries for the period 1990-2003. The measure of linguistic distance used in their study was based on the work of Dyen et al. (1992): Dyen et al. (1992) constructed a measure of the distance between Indo-European languages based on the proximity of 200 words from each language (See Box 3). This variable was included in the estimating equations along with a common language variable. Unlike the research of Belot and Hatton (2012), both the common language and linguistic distance variables were statistically

[5] Belot and Hatton (2012) noted that they also estimated equations that included a dichotomous common language variable, and that this common language variable was insignificant in the presence of the linguistic proximity variable.

significant, although the impact of sharing a common language dropped by over one-quarter when the measure of linguistic distance was included in the estimating equation.

Box 3: The Dyen Lexicostatistical Percentage Approach

Comparative lexicostatistics is the study of historical relations among speech varieties belonging to the same language family through a quantitative study of cognation among their vocabularies. The lexicostatistical percentage approach is the oldest and most widely-used lexicostatistical approach. As explained by Dyen et al. (1992), their application of this approach has four phases. First, they worked with the 200 meanings that had been proposed by Swadesh, and developed phonetic representations (forms) of the words with these particular meanings for the chosen languages (see Swadesh (1952) for an earlier compilation). Then the cognation among the forms in two languages was established through expert judgment. Cognation requires that the forms have descended in unbroken lines from a common ancestor in the same language family. Consideration of the number of cognate forms from the list of 200 meanings gives rise to the so-called lexicostatistical percentage. For example, the value when German and English are compared is 57.8 percent. The value for the French-English comparison is 23.6 percent. In other words, German and English are more similar than French and English. The final phase of work by Dyen et al. (1992) involved the categorization of the languages into various groups. It is to be noted that the Dyen et al. work covers only Indo-European languages. An example of Dyen et al.'s (1992) numbers is given below.

Dyen Matrix of Linguistic Distances (higher values mean smaller distance)

	Languages							
Languages	Italian	French	Spanish	German	Dutch	Danish	English	Greek
Italian	1.000	0.803	0.788	0.265	0.260	0.263	0.247	0.178
French	0.803	1.000	0.734	0.244	0.244	0.241	0.236	0.157
Spanish	0.788	0.734	1.000	0.253	0.258	0.250	0.240	0.167
German	0.265	0.244	0.253	1.000	0.838	0.707	0.578	0.188
Dutch	0.260	0.244	0.258	0.838	1.000	0.663	0.608	0.188
Danish	0.263	0.241	0.250	0.707	0.663	1.000	0.593	0.183
English	0.247	0.236	0.240	0.578	0.608	0.593	1.000	0.162
Greek	0.178	0.157	0.167	0.188	0.188	0.83	0.162	1.000

Examples of economics studies using these data are Ginsburgh, Ortuño-Ortín and Weber (2005), Belot and Ederveen (2012) and Adresa and Pytlikova (2012).

Source: Dyen, I., Kruskal J.B. and Black, P., (1992). An IndoEuropean Classification: A Lexicostatistical Experiment, *Transactions of the American Philosophical Society*, New Series 82(5).

An idea of the relative importance of the Belot and Ederveen (2012) measure of linguistic distance for migrant flows can be found from the effects that these authors computed for one standard deviation increases in the various explanatory variables. They report that 'Our regression results imply that an increase in linguistic distance with one standard deviation lowers the migration flow with 56%...This effect is about 50% higher than the effect of raising GDP per capita in the destination country by one standard deviation and much more than a change of one standard deviation in unemployment rates' (Belot and Ederveen, 2012, p.1096). The importance of the linguistic distance measure in the analysis of migration flows was found in the many tests of robustness these authors conducted.

The final study in our review is by Adsera and Pytlikova (2012). They cover immigration flows in 30 OECD countries from 233 source countries, for the years 1980-2009. Their research is important in the study of the links between linguistic distance and the destination choice of immigrants because they used many measures of linguistic distance or linguistic proximity. Their preferred measure was based on the *Ethnologue: Language of the World*, and ranges from zero to one according to the number of levels of the language family tree shared by the destination and source country languages (See Box 2). Belot and Hatton (2012) also use this type of measure. In addition, in tests of robustness, Adsera and Pytlikova (2012) use both a measure based on Dyen et al. (1992) (see also Belot and Ederveen (2012)) and a measure based on the Levenshtein linguistic distance approach produced by the Max Planck Institute for Evolutionary Anthropology. See Box 4.

Adsera and Pytlikova (2012) report that their preferred measure of linguistic proximity is a statistically significant determinant of migration flows.[6] This holds in both bivariate and various multivariate models, including those that take account of both the stock of migrants in the destination and the flow of migrants between countries in the previous period, as well as in models that include destination and origin country fixed effects. Moreover, the finding that linguistic proximity is an important determinant of migration flows was robust with respect to the use of the two alternative measures of linguistic proximity (based on Dyen et al. (1992) —see Box 3; and the Max Planck Institute for Evolutionary Anthropology—see Box 4). It was also robust with respect to the choice of language to use in the construction of the proximity measure (main official language, any official language, and the major language, where major was defined as that which was used most extensively). It was also robust with respect to when the effects are estimated separately for English-speaking countries and for non-English-speaking destination countries. The estimated impacts were, however, stronger for non-English-speaking destination countries. Adsera and Pytlikova (2012, p.25) argue that 'The likely higher proficiency of the average migrant in English rather than in other languages may diminish the relevance of the linguistic proximity indicators to English speaking destinations'. This greater proficiency in English is likely due to English having become the international language of science, technology, and business.

[6] Adsera and Pytlikova (2012) also used measures of the diversity of languages in both the country of origin and the country of destination.

Box 4: The Max Planck Institute for Evolutionary Anthropology measure of Levenshtein linguistic distance

The Max Planck Institute for Evolutionary Anthropology in Germany has used a 'lexicostatistical" approach to develop a measure of linguistic distance, or more precisely the Levenshtein distance, using an algorithm that compares pronunciation and vocabulary of language pairs. This procedure is based on the Automatic Similarity Judgment Program (ASJP). The starting point for this approach is a small Swadesh list (see Box 3) of 40 words that describe common things and environments. These words are then expressed in a special phonetic transcription known as the ASJP code. This code uses 41 characters on a standard QWERTY keyboard to represent the common sounds in human communication. Then the number of additions or subtractions of characters (or sounds) required to transform a word in one language into the same word in another language is computed, using the ASJP. Isphording and Otten (2012, p.7) offer the following illustration:

> …to transfer the phonetic transcription of the English word *you*, transcribed as *yu*, into the transcription of the respective German word *du*, one simply has to substitute the first consonant. But to transfer *manunt3n*, which is the transcription of mountain, into *bErk*, which is the transcription of the German *Berg*, one has to remove or substitute each 8 consonants and vowels, respectively.

This evaluation is then adjusted to account for differences in world length and the potential similarities in phonetic inventories that might lead to similarity by chance to give the Levenshtein distance measure. Larger values thus indicate languages that are further apart. Some examples of this measure of linguistic distance are provided below.

Examples of Closest and Furthest Language Pairs with Respect of the Levenshtein Distance Measure

	Closest		Furthest
Language	Distance	Language	Distance
Distance to English			
Afrikaans	62.08	Vietnamese	104.06
Dutch	63.22	Turkmen	103.84
Norwegian	64.12	Hakka (China)	103.10
Distance to Spanish			
Galician	54.82	Wolof (Senegal)	103.02
Italian	56.51	Igbo Onitsha (Nigeria)	102.84
Portuguese	64.21	Ewondo (Cameroon)	101.87

Source: Extracted from Table 1 in Isphording and Otten (2012).

This data source has been used in economic research by Isphording and Otten (2011)(2012) and Adsera and Pytlikova (2012).

Table 1 summarizes the evidence on the links between language background and the choice of destination among immigrants. It shows that language background matters to destination choice, and that stronger empirical results emerge in studies that use measures of linguistic distance than in studies that employ simple dichotomous variables to reflect a common language between pairs of countries. In other words, it is not just the knowledge of a destination dominant language that matters, but the ease with which an immigrant can learn the destination dominant language is also very important.

Table 1: Overview of Studies into the Links Between Language Background and Destination Choice of Immigrants

Authors	Countries studied and time period	Language variable	Does language matter to destination choice?
Clark, Hatton and Williamson (2007)	Flows to the US from 81 source countries, 1971-1998	English dummy	Yes
Karemera, Oguledo and Davis (2000)	Flows to US and Canada, 1976-1986	English/English or French dummy	No
Pedersen, Pytlikova and Smith (2008)	Flows to 22 OECD countries from 129 source countries, 1990-2000	Common language dummy	Yes
Mayda (2010)	Flows to 14 OECD countries from 79 sending countries, 1980-1995	Common language dummy	No
Ortega and Giovanni (2009)	Flows to 14 OECD countries from 73 source countries, 1980-2005	Common language dummy	No
Grogger and Hanson (2011)	Stock of immigrants in 15 high-income OECD countries in 2000	Common language dummy	Yes
Beine, Docquier and Özden (2011)	Change in stocks between 1990 and 2000, for 30 OECD countries, with 195 source countries	Common language dummy	Yes
Belot and Hatton (2012)	Stock of immigrants from 70 source countries in 21 OECD countries, 2000-2001	Based on the Ethnologue (see Box 2)	Yes
Belot and Ederveen (2012)	Flow of immigrants to 22 OECD countries, 1990-2003	Based on Dyen et al. (1992) (see Box 3)	Yes

Adsera and Pytlikova (2012)	Flows of immigrants to 30 OECD countries from 233 source countries, 1980-2009	Based on the Ethnologue (see Text Box 2), Dyen et al. (1992) (see Box 3) and the Levenshtein distance approach (see Box 4)	Yes

3.2 *Location choice within a country*

The seminal study on the location choice of immigrants within a country is Bartel (1989) which used 1980 Census data.[7] This covered the US. The main set of results in this paper was from a conditional logit model, estimated separately for Asians, Hispanics and Europeans. The explanatory variables used in Bartel's (1989) model included characteristics of the areas within which the immigrants lived, such as the unemployment rate, average wage, and the distance of the location from the immigrant's country of origin. The main such variable, however, was the ethnic concentration measure, defined as the percentage of a specific ethnic group that resided in the particular location. The ethnic concentration variable was a highly significant determinant of location choice for each ethnic group, and within each ethnic group, for the three arrival groups considered. This effect tended to be weaker among the more educated immigrants. Using the Census information on place of residence five years ago, Bartel (1989) examined internal migration patterns. Immigrants were reported to be more likely to change locations than the native born, and much of this movement was associated with an increase in the geographical concentration of the ethnic group. While this pioneering study did not examine the role of

[7] There is no difference in the theory of migration between internal and international migration. Institutional factors, such as regulations, may differ.

language background per se, the apparent links between ethnic concentrations and language backgrounds were developed in subsequent research.[8]

The approach of Bartel (1989) has been developed by Jaeger (2000), among others. Jaeger (2000) used data from the Immigration and Naturalization Service on immigrants admitted to the US during the 1990-1991 fiscal year, combined with data from the 1980 and 1990 censuses. Bartel's (1989) ethnic concentration variable was expanded to consider the immigrant's region of birth, and the share of immigrants who speak a language other than English that is spoken in the immigrant's country of birth. These location characteristics were considered in conjunction with the share of the population in the location that was born abroad. Jaeger (2000, p.15) reported that 'Region-of-birth concentrations are about 3 times as important in determining location as language and foreign-born shares'. This relative importance held for all visa types other than for the small group of diversity visa immigrants. It also carried over to the analysis of the location choice of the foreign born in the US who received an adjustment of state to become permanent resident aliens. Nevertheless, this research showed that language background matters for the location choice of immigrants within a country.

The research by Bauer, Epstein and Gang (2005) is focussed on the links between the location choice of immigrants from Mexico in the US and their English language proficiency. Specifically, they ask whether the choice of the size of enclave community in which to settle is affected by English language proficiency. The analyses were based on data from the Mexican

[8] Zavodny (1999) reported that the fraction of the state population that is foreign born was a highly significant determinant of recent immigrants' location choice for all admission categories considered in her study (family, employment, Immigration Reform and Control Act of 1986 (IRCA) conversions to legal status, and refugee/asylee conversions to legal status), and for all country groups examined (Chinese, Dominican Republic, Mexico, Philippines, Vietnam).

Migration Project, which collected information on migrants to the US from their communities of origin in Mexico. The main variables of interest in this study are the proportion of the total population in a particular US location that was from Mexico, and the English speaking proficiency of the immigrants (can speak and understand English; can understand but not speak English; can neither speak nor understand English). The authors report that ethnic enclave effects are strongest among those who can neither speak nor understand English, and are weakest among those who can both speak and understand English. These results were broadly the same for both first-time movers and for repeat movers. Chiswick and Miller (2005b) show that residence in an enclave community reduces an immigrant's own destination language proficiency. Combining these results, Bauer et al. (2005, p.660) concluded '…enclaves are a potential source for a "language trap"; they attract poor proficiency English speakers and sustain their poor abilities'. The findings of Bauer et al. (2005) were robust with respect to alternative definitions of an ethnic enclave.

Turning to the Canadian literature, which is of interest due to the language divide between Quebec and the rest of Canada, there are several relevant studies. Hou (2005) contains a detailed analysis of Census of Canada data over the period 1981 to 2001. The tabulations in this study revealed that as many as 90 percent of immigrants from Haiti, where French is an official language and Haitian creole is close to French, settle in Montreal. The disproportionate representation (only about 11 percent of all immigrants in Canada live in Montreal) was maintained when the location of immigrants was examined after 11 to 15 years of residence in Canada.

McDonald (2003)(2004) used data from the Statistics Canada Censuses for 1986, 1991 and 1996 to estimate conditional logit models of the initial location decision of immigrants in Canada. He estimated models with an aggregate-level ethnic concentration measure, and with measures based on the age, educational attainment, and official language skills of the immigrant population. The impacts of these ethnic concentration measures were allowed to vary between those who usually spoke English or French at home, and those who did not. It was shown that immigrants who usually spoke a language other than English or French at home were more likely to settle in a linguistic enclave community than those who spoke English or French at home. The ethnic enclave variable (based on those who speak neither of the official languages at home) was also a statistically significant determinant of initial location choice. The effect of this enclave influence also differed between those who usually spoke English or French at home and those who did not. Hence, both the relative concentration of immigrants from the same ethnic group, as well as the characteristics of these ethnic enclaves, affect the location decisions of recent immigrants. The importance of language background was made clear in the simulation that McDonald (2004) presented for immigrants from the non-Arab countries of Africa. Where the immigrants spoke English at home, the distribution across regions was heavily concentrated in English-Canada. But where the immigrants spoke French at home the distribution across regions was heavily concentrated in Montreal and the rest of Quebec.

Thus, although the literature on the importance of language background to immigrants' location decisions within a country is sparser than that relating to the choice of destination country, the evidence suggests that language skills matter to the way migrants distribute themselves across regions within a country. This evidence is summarized in Table 2.

Table 2: Overview of Studies into the Links Between Language Background and Choice of Location Within a Country

Authors	Country studied and time period	Language variable	Does language matter to location choice?
Jaeger (2000)	US, Immigration and Naturalization Service on immigrants admitted during 1990-1991	Origin language concentration	Yes, for almost all visa types
Bauer, Epstein and Gang (2005)	US, Mexican Migration Project, annual data collection that commenced in 1987	Immigrants from Mexico concentration	Yes, for first-time movers and repeat movers
Hou (2005)	Canada, Census of Canada, 1981 to 2001	Distribution of immigrants by origin countries which differ in language background	Yes
McDonald (2003)(2004)	Canada, Census of Canada, 1986 to 1996	Ethnic concentration, with focus on official language skills	Yes

4. Determinants of Language Proficiency

Research on the determinants of dominant language proficiency among immigrants from a different linguistic background than the destination has focused on three concepts represented by the three 'Es': Exposure to the host country language, Efficiency in learning a new language and Economic incentives for learning the new language (Chiswick 1991, Chiswick and Miller, 1995, 2007a). These are conceptual variables, but empirical research requires finding measurable dimensions. Here we review the empirical literature on dominant language proficiency among immigrants using this three 'Es' framework. From humble beginnings in the 1980s, this empirical literature has grown enormously. Thus, given the volume of studies, the review that follows is illustrative rather than exhaustive.

An important methodological matter that needs clarification is the measure for the language variable. Two broad measures have been used in the literature. The first is dominant language usage, typically termed 'language shift' from origin language to destination dominant language. For example, see Veltman (1983), Grenier (1984) and McAllister (1986). An alternative terminology is 'mother tongue retention' (Chiswick and Miller, 2008b). Dominant language usage in preference to the language of the country of origin is generally viewed as a measure of cultural assimilation. The second measure is dominant language proficiency, that is, how well the person can speak (or, in some studies, understand, write or read) the main language of the destination country. This is the human capital skill that researchers focus on when studying labor market outcomes. Dominant language proficiency is focused on in this review, although several key findings from studies of dominant language use are mentioned.

4.1 *Exposure*

Much of destination language learning among immigrants comes from exposure to the destination language. Exposure can be thought of as having two dimensions, that is, exposure in the origin and exposure after migration.

The data sets used to study the determinants of immigrant's destination language skills generally indicate the country of origin, but provide no direct information on pre-immigration language learning.[9] When conducting research on English-speaking destinations for immigrants from non-English speaking origins, a proxy measure for pre-migration exposure to English is

[9] In analysis of the determinants of German language skills among immigrants in Germany with Italian, Spanish, Yugoslavian, Turkish or Greek nationality, Dustmann (1994) assumed that knowledge of the German language was non-existent at the time of migration.

whether the origin was a former colony or dependency of either the United Kingdom or the United States. Immigrants in the UK or the US from former colonies (e.g., Nigeria, India or the Philippines) are found to be more proficient in English than are immigrants from other (non-English-speaking) countries that were not dependencies of the UK or the US (e.g., Thailand or Algeria), other variables being the same (Chiswick and Miller, 2001, 2007a). Similarly, immigrants in Spain from former colonies were reported by Isphording and Otten (2012) to be more proficient in Spanish than immigrants from other countries.

Another way of capturing pre-immigration exposure is to categorize countries according to whether the dominant language of the destination country is an official language or the dominant language of the country of origin, using sources such as the Ethnologue: Languages of the World (see Box 2). Espenshade and Fu (1997), for example, reported that the language spoken in the country of origin is an important determinant of English language skills among non-native English-speaking immigrants in the US. In analyses based on a binational source of data on Mexico-US migrants, Espinosa and Massey (1997) reported that the English proficiency is higher among migrants from communities (in Mexico) with greater proportions of adult men with US migrant experience, a variable which is argued to capture the 'degree of contact with U.S. culture within the respondent's community' (Espinosa and Massey, 1997, p.37).

Thus, both the broad indicators provided by country of origin groupings, and the detailed information on origin country exposure where available, present a consistent set of evidence that pre-immigration exposure matters. Pre-immigration exposure has also been found to be an important determinant of proficiency in the dominant language of the destination country in the few studies that have been able to include direct measures. Raijman (2013), for example, studied

the proficiency in Hebrew among Jewish South African immigrants in Israel, and reported that the level of Hebrew proficiency before arrival (typically acquired through attendance at Jewish schools and participation in synagogue activities and youth movements) was a highly significant and quantitatively important determinant of post-arrival proficiency.

The most important aspect of exposure to the destination language occurs after migration. Exposure in the destination can be decomposed into time units of exposure and the intensity of exposure per unit of time. Most data that identify the foreign-born members of the population ask the respondents when they came to the destination. From this, a variable for duration or 'years since migration' can be computed. Duration has a very large positive and a highly statistically significant impact on destination language proficiency, but the effect is not linear. Rather, proficiency increases rapidly in the early years, but it increases at a decreasing rate; hence after a period of time a longer duration in the destination has a much smaller positive impact (Chiswick and Miller, 2001, 2007a, 2008b; Espenshade and Fu, 1997; Isphording and Otten, 2011, 2102). Grenier (1984) reported a similar pattern in his study of shifts from Spanish to English as the usual language among Hispanics in the US.

This time pattern for destination language proficiency is likely to be due to incentives for investment in language skills. For the following three reasons, an immigrant has the incentive to make greater investments shortly after arrival rather than delaying investments: to take advantage sooner of the benefits of increased proficiency, to make the investments when the value of the immigrants' time (destination wage rate) is lower, and to have a longer expected future duration in the destination.

Duration may affect language proficiency because a longer actual duration increases the

amount of exposure to and practice using the destination language. It is found that interrupted stays, that is, when immigrants move back and forth (sojourners), reduce their language proficiency (Chiswick and Miller, 2001, 2007a, 2008b; Isphording and Otten, 2102). The expectation of an interrupted stay reduces the incentive to invest, implicitly if not explicitly, in language learning, and the destination language skills tend to depreciate during long periods of absence from the destination.

Moreover, those in the destination who report that they expect to return to their origin are also less proficient, other variables being the same (Chiswick and Miller, 2006). This might arise from negative selectivity in return migration, that those having a more difficult adjustment to the new country are more inclined to leave to return to the origin or to go to a third country. Or, it might reflect the reduced incentive to invest in destination language skills if the expected future duration (i.e., the payoff period) is short.[10]

Espinosa and Massey (1997) were able to include very detailed information on individual's migration history in their study of Mexico-US migrants. Included are variables for the period of first entry, the period of last entry, the total time the individual had spent in the US since the first entry, the proportion of time spent in the US since the first entry, the average number of trips taken per year between the first and most recent visits, and the duration of the most recent trip. They report that English language skills increase with each of the latter four variables.

The intensity of exposure per unit of time in the destination is usually more difficult to

[10] Selective emigration could be associated with biased cross-sectional estimates of the coefficients in the models of dominant language proficiency. Espinosa and Massey (1997), however, on the basis of analysis of data on Mexico-US migrants collected in both Mexico and the US, conclude (p.44) 'Our analysis reaffirms most of the findings established by prior studies of linguistic assimilation, lending some confidence to the belief that they are not simply artefacts of sample selection or omitted variable bias'.

measure. A few studies have included information on whether immigrants enrolled in formal classes of instruction in the destination language. Raijman (2013), for instance, related Hebrew proficiency of South African immigrants in Israel to whether they studied Hebrew in Israel for more than six months (i.e., though attending an uplan, which is an institute or school designed to teach adult immigrants basic Hebrew skills through an intensive course of instruction). Immigrants in this situation had higher levels of Hebrew proficiency than immigrants who studied for shorter periods or who did not enroll in a formal program of instruction in Hebrew. More often in the research, however, the focus is on the environment in which one lives, comprising both the area and the family. In terms of the area, it is useful to have a proxy measure of the ability to avoid using the destination language. Various measures have been used in the different studies, though the construct used most often is a 'minority language concentration' measure. This is typically constructed as the percentage of the population, including the native born and the foreign born, in the area (defined by the state/province, region or metropolitan area) where the respondent lives, who speak the same non-destination language as the respondent. For example, the concentration measure for an Italian speaker living in Chicago would be the proportion of the population of Chicago who speak Italian. In other instances, newspapers (Australia) or radio broadcasting (United States) in the language of origin have been used either as a substitute for, or in addition to, the minority language concentration measure.[11] The effects on language proficiency of these area-based minority language concentration measures are quite strong. Destination language proficiency is significantly lower among individuals who have greater ease in avoiding using the destination language by living in a linguistic enclave area

[11] To mitigate the problem of endogeneity, the instrumental variables technique was used to obtain predicted values for the newspapers and broadcasts (Chiswick and Miller, 1995)

(Chiswick, 1998; Chiswick and Miller, 2007a, 2008b; Espenshade and Fu, 1997; Isphording and Otten, 2012; Lazear, 1999; Warman, 2007). Similarly, destination language use is less likely in areas with a geographical environment that favors interactions in the origin country language (Grenier, 1984). Where direct measures of social contact have been available (McAllister, 1986), such as the presence of 'close friends from the country of origin', the finding that such contacts reduce dominant language proficiency reinforces those based on the more general characteristics of the area of residence. Similarly, Espinosa and Massey (1997) report that English proficiency is higher among immigrants from Mexico in the US who have more extensive contacts with members of US racial and ethnic groups.

A key role in language learning is played by the family or household in the destination in which the immigrant lives. Both the spouse, if married, and the children matter. Those who married their current spouse before immigrating are likely to be married to someone with the same language background. They are more likely to speak that language to each other at home, thereby limiting opportunities for practicing the destination language at home.[12] On the other hand, those who marry after immigration are more likely to marry someone proficient in the destination language, perhaps because of their own proficiency, and are more likely to practice the destination language. Where the data permit a study of this issue, it is found that, other measured variables being the same, the most proficient are those who married after migration, followed by those who are not married, with those who married their current spouse before migration being the least proficient (Chiswick and Miller, 2005b, 2007a, 2008b; Chiswick et al. 2005a, 2005b; Dustmann, 1994). Grenier (1984) reported that, compared to the non-married,

[12] Akresh (2007) reports that among those immigrants for whom English was not a native language, English was most likely to be used at work, and least likely to be used with one's spouse.

English language use at home was more likely among married Hispanics whose spouse was non-Hispanic, and less likely among Hispanics who were married to a Hispanic. The more direct evidence reported by Dustmann (1994) adds to this: based on analysis of immigrants in Germany, he reported that proficiency in German was higher where the partner has good German speaking skills. Similarly, Espenshade and Fu (1997) show that English skills are lowest where the spouse is from the same non-English language dominant country, and highest when the spouse is from any English language dominant country.

Children can have offsetting effects on their parents' proficiency (Chiswick, 1998; Chiswick and Miller, 2007a, 2008b; Chiswick et al., 2005a, 2005b). For example, children can serve wittingly, or unwittingly, as 'teachers'. Whether they themselves are immigrants or not, children learn the destination language quickly because of their youth and because of their exposure to the destination language in school. They can, therefore, bring it home to their parents.

Yet, the presence of children can also have negative effects on their parents' proficiency. Parents may speak the language of the origin at home to transmit the origin culture to their children, in part so that their children are able to communicate with the grandparents and other relatives who did not migrate or who migrated but lack proficiency in the destination language. Children may also serve as translators for immigrant parents. The translator role may be more effective in consumption activities and in dealings with the government bureaucracy and the educational and health care systems than in the workplace. Finally, children tend to reduce the labor supply of their mothers who stay at home to provide childcare. To the extent that adults invest in improving their language skills in anticipation of labor market activities, and benefit

from doing so, and to the extent that practice using the destination language at work enhances proficiency, children would tend to be associated with lower proficiency among their mothers.

Taken as a whole, the four hypotheses regarding children suggest an ambiguous effect on their parents' proficiency, but due to the latter two, their effect would be less positive or more negative for their mothers than their fathers. Empirically, this is in fact what is found. Where there is no clear effect of children on their father's proficiency, in the same data, it is less positive or more negative for their mothers (Chiswick and Miller, 2007a; Chiswick and Repetto, 2001; Chiswick, et al. 2005a, 2005b; Grenier, 1984). Where there is a positive, albeit small, effect of children on their father's proficiency, the effect for mothers is statistically insignificant (Dustmann, 1994).

There is language learning in the home. Research has shown that the proficiency of one family member is positively associated with that of other family members (Chiswick et al. 2005b). The children's proficiency is more highly correlated with that of their mothers than with that of their fathers. This makes sense since mothers are generally more directly involved in the raising of their children than are the fathers. Similarly, Espinosa and Massey (1997) report that the English proficiency of migrants from Mexico in the US was higher where they had siblings who were US migrants, and where they had children in US schools.

As a result, particularly due to a weaker attachment to the labor force, immigrant women with children have a lower level of destination language proficiency than do men and than do women without children (Chiswick et al., 2005a, 2005b; Stevens, 1986).

4.2 *Efficiency*

The second ‘E’, efficiency, refers to the ability to convert exposure into language learning. Age at migration is an important efficiency variable. Because of the greater plasticity of the brain, which decreases with age, language learning decreases significantly with a greater age at migration (Long, 1990). There is a debate in the linguistics literature regarding the ‘critical period hypothesis’, whether there is an age beyond which an immigrant’s learning a second language, that is, the destination language, becomes much more difficult. The chart in Figure 1, based on data on self-reported speaking proficiency at home from the US 2000 Census of Population for foreign-born males and females, shows the negative relation between proficiency and age at migration. Note that the path for this measure of proficiency is remarkably similar for men and women, a pattern which has also been reported by Dustmann (1994).

The Figure 1 data do not suggest any particular critical age at migration for speaking proficiency (Chiswick and Miller, 2008a). This does not rule out a critical period for other dimensions of proficiency, such as retaining an accent. This chart does, however, convey a main finding from the empirical research on the dominant language skills among immigrants, that age at migration is an important determinant (Chiswick and Miller, 2008a; Espenshade and Fu, 1997; Grenier, 1984; Isphording and Otten, 2012).

Figure 1: Effects of age at migration on English language proficiency among immigrants from non-English speaking countries, by gender (United States, 2000)

FIGURE 1 HERE

Source: US Census of Population, 2000, Census Public Use Microdata Sample – Sample of the Population.

Education is considered to be another efficiency variable. Other variables being the same, empirically immigrants with more schooling are more proficient in the destination language. This could arise because those with higher levels of schooling are more efficient learners, either inherently (higher ability people get more schooling), or because they acquire learning skills in school. To some extent, this effect for immigrants in the United States, Canada and Australia might be due to being exposed to English as the future immigrants advance up the educational system in the origin (Chiswick and Miller, 1995, 2007a, 2008b; Espenshade and Fu, 1997). It should be noted, however, that this is not likely to be a dominant factor since there is a similar relationship between schooling and Hebrew language skills among immigrants in Israel and most immigrants to Israel arrive without proficiency in Hebrew (Chiswick, 1998; Chiswick and Repetto, 2001). Similarly, Dustmann's (1994) study of immigrants in Germany, where it was noted that German was not the first foreign language learned at school in the origin countries, shows that years of schooling was a significant determinant of Germany speaking skills. Isphording and Otten's (2011)(2102) analyses for Germany also exhibit a strong influence of years of schooling on dominant language proficiency. Notably, where attempts are made to distinguish years of schooling undertaken in the country of origin, and years of schooling undertaken after arrival in the destination country, the latter variable has been shown to have the more important impact on destination language skills (see Evans (1986) for Australia,

Espenshade and Fu (1997) and McManus et al. (1983) for the US and Dustmann (1994) for Germany).

The efficiency with which an immigrant can learn the dominant language of the destination country could also vary with their origin country language skills. Dustmann (1994) and Isphording and Otten (2011) both report that, in Germany, immigrants with very good ability in writing the home country language were more likely to be fluent in German, whereas those who were classified as illiterate in the home country language were less likely to be fluent in German.

Some languages share many similarities with English (e.g., Dutch), while others are very different (e.g., Korean) and hence make it more difficult to learn English. Language trees have been used by linguists to map out the evolution or historic relations among languages (Crystal, 1987; Lewis, 2009). But what is needed for a statistical analysis of the determinants of the effect of 'linguistic distance' on English language proficiency is a quantitative measure of the difficulties that non-English speakers have in learning English.[13] One such measure has been developed and tested using an index of the difficulty that Americans have learning other languages and the assumption of symmetry (i.e., if the Americans have difficulty learning Korean, then the Korean speakers would have difficulty learning English) (Chiswick and Miller, 2005a); see Box 5. This measure of linguistic distance has been shown to be important for understanding the English language proficiency of immigrants in the United States, Canada and Australia (Chiswick and Miller, 2005a). In principle, the methodology could be applied to develop measures of linguistic distance from other languages.

[13] See McCloskey (1998, pp.104-106) for an interesting perspective. The heading of McCloskey's discussion is 'In Like Fashion, Rhetorical Standards Are Necessary in Linguistics to Measure the Similarity of Languages'.

Recently, Isphording and Otten (2011)(2012) have proposed the use of a measure derived using a Levenshtein distance approach: see Box 4. They note that, compared to the measure of Chiswick and Miller (2005a) which has been developed for only one destination country language, English, the Levenshtein distance measure can be readily computed as a continuous variable for any pair of host and home country languages, and does not rely on a symmetry assumption.[14] These authors have applied this measure of linguistic distance in analyses of the dominant language acquisition of immigrants in the US, Germany and Spain. The results indicate a strong, significantly negative, effect of linguistic distance on immigrant language skills. For the analyses undertaken for the US, the authors also compared the findings obtained using the Levenshtein distance measure with the test-score-based measure of linguistic distance of Chiswick and Miller (2005a). It is shown that these two linguistic measures render qualitatively comparable outcomes.

[14] Isphording and Otten (2012) argue that the linguistic distance measure of Chiswick and Miller (2005a) could be biased by incentives for learning a foreign language, while the measure based on Levenshtein distance should be devoid of any such bias.

Box 5: The Chiswick and Miller measure of linguistic distance

Chiswick and Miller (2005a) construct a scalar measure of the distance between English and a myriad of other (non-native American) languages. It is computed using a set of language scores, reported by Hart-Gonzalez and Lindemann (1993), which measure the achievements in speaking proficiency by English-speaking Americans at the US Department of State, School of Language Studies. Specifically, Hart-Gonzalez and Lindemann (1993) report the level of proficiency in a language, using the average exam score after 24 weeks of lessons. These measures are an index of the difficulty native English-speakers have learning the language. It is assumed that there is linguistic symmetry: The more distant a language is from English, the more difficult it is for speakers of that language to learn English. These scores range from 1.00 (hardest to learn) to 3.00 (easiest to learn). Examples are provided below.

Linguistic Score	Illustrative Languages
1.0	Korean, Japanese
1.5	Vietnamese, Arabic
2.0	Polish, Indonesia
2.5	Portuguese, Italian
3.0	Norwegian, Swedish

This linguistic distance score has been criticized by Isphording and Otten (2012, p.6) because 'It has to be assumed that the difficulty of U.S. citizens to learn a particular foreign language is symmetric to the difficulty of foreigners to learn English. Further, it has to be assumed that the average test score is not influenced by other language-specific sources'.

This measure of linguistic distance has been applied in analyses of earnings and English fluency by Chiswick and Miller (2005a) and Isphording and Otten (2102), and in analyses of the determinants of bilateral trade by Hutchinson (2005).

Source: Chiswick, B.R. and Miller, P.W., (2005a). 'Linguistic Distance: A Quantitative Measure of the Distance between English and Other Languages', *Journal of Multilingual and Multicultural Development* 26(1), pp.1-16.

Another efficiency variable is the motive for migrating. Three broad categories can be distinguished: employment migrants, refugees and family migrants. The employment migrants are most favorably selected for labor market success in the destination since this is their primary

motivation. The refugees, on the other hand, include many who would not have moved except for the political, ethnic or religious problems they confronted in their origin. As a result, they tend to have the lowest degree of selectivity for success in the destination, and would be expected to have the least transferability of their skills (Chiswick and Miller, 2007a, 2008b). Lawyers and judges, for example, are very rare among employment-motivated migrants because their skills are not readily transferable across countries, but they are not uncommon among refugees. Family migrants fall in between employment migrants and refugees as they are attracted by economic opportunities in the destination as well as family ties, but are not responding to the same forces as refugees.

In some data, the motive for migration or the visa used to gain entry can be identified. When this has been possible, it is found that the employment-motivated immigrants have the highest level of destination language proficiency, followed by the family-based immigrants, with the refugees showing the lowest proficiency (Chiswick and Miller, 2006, 2007a). Espinosa and Massey (1997) examined the independent effect of legal US residence in their study of immigrants from Mexico in the US. They reported (Espinosa and Massey, 1997, p.44) that 'Migrants who lack documents actually spoke and understood significantly more English at the time of their most recent trip than did those who already had received legal U.S. residence'. This finding is counter-intuitive, and it is possible that it is attributable to the porous nature of the Mexico-US border.

Immigration policy can also affect the destination language proficiency of immigrants. Some countries (e.g., Canada and Australia) give explicit preference to applicants for immigration who can demonstrate proficiency in the destination language. When Australia

increased the language proficiency requirements for employment-based independent immigrant visas, but not for other visas, the English language proficiency of the skills-tested migrants increased, with no significant change for the other groups (Chiswick and Miller, 2006).

4.3 *Economic Incentives*

The economic incentives for acquiring destination language proficiency also play an important role. The returns to becoming proficient are greater when the expected duration in the destination is longer, whether as a worker or as a consumer. Various proxy measures of the expected future length of stay in the destination have been employed, depending on the data available. These include self-reported expectations of the duration of stay (Australia, Germany), re-migration rates of immigrants from the respondent's country of origin (United States and Israel), and the distance from the origin (United States, Australia and Canada), since return migration propensities decline with distance. Regardless of the measure, the longer the expected duration of stay, the greater is the investment in destination language proficiency (Chiswick and Miller, 2006, 2007a, 2007b, 2008b; Dustmann, 1999; Isphording and Otten, 2011).

The most problematic aspect of the research on the determinants of destination language skills is estimating the impact on proficiency of the expected increase in earnings from becoming more proficient, that is, using the individual's expected increase in earnings as an explanatory variable. Data are not available for this on an individual basis. Those with higher levels of skills, for example, professionals as distinct from laborers, tend to gain relatively more in earnings from proficiency (Chiswick and Miller, 2003, 2007a). If so, the education variable would reflect some of this effect. Dustmann (1994) addresses the incentive to learn issue through the inclusion of a

variable for whether the immigrant had ever worked in an analysis of German speaking fluency estimated for females. This variable had a sizeable effect, which Dustmann (1994, p.141) argued could '…indicate that those who do not intend to ever participate in the labor market have lower incentives to learn the German language'.

The findings reported here for the determinants of destination language proficiency among immigrants are remarkably robust across types of data (census or survey, cross-sectional or longitudinal), countries of destination (United States, Canada, Australia, Germany and Israel) and countries of origin (Chiswick and Miller, 2007a).[15] Findings in relation to writing skills are, however, distinguished from those for speaking skills by the fact that variables which represent contacts with the host country population are not statistically significant (see Dustmann, 1994).

The evidence reviewed above is important in terms of showing that the patterns in most data sets analyzed are highly consistent with the three 'Es' model of the development of destination country dominant language skills among immigrants. Table 3 provides an overview of the empirical support for the model.

[15] Charette and Meng (1994) reported that the findings for the determination of language proficiency may not be robust to whether a self-assessed or objective measure is used. This issue has not attracted much attention, most likely because it does not appear to be empirically relevant in research into labor market outcomes (Charette and Meng, 1994). Chiswick, Lee and Miller (2003), however, argued that both sets of measures contain useful information.

Table 3: Overview of Empirical Support for the Three 'Es' Model of Destination Language Proficiency

Concept, Empirical Counterparts and Expected Sign	Selected Studies Reporting Supportive Evidence
A. *Exposure* (+)	
A.1 Pre-immigration exposure	
• Former colony or dependency (+)	Chiswick and Miller (2001)(2007a), Isphording and Otten (2012)
• Official language/dominant language status of origin country (+)	Espenshade and Fu (1997)
• Pre-immigration contact with dominant language culture (+)	Espinosa and Massey (1997)
A.2 Post-immigration exposure (+)	
A.2.a Time units of exposure (+)	
• Duration of residence in destination country	Chiswick and Miller (2001)(2007a), Espenshade and Fu (1997), Isphording and Otten (2011)(2012)
• Interrupted stay (-)	Chiswick and Miller (2001)(2007a)(2008b), Isphording and Otten (2012)
A.2.b Intensity of exposure (+)	
• Minority language concentration (-)	Chiswick and Miller (2001)(2007a), Espenshade and Fu (1997), Lazear (1999), Isphording and Otten (2012)
• Direct measures of contact with members of the destination (+) or origin countries (-)	McAllister (1986), Espenshade and Fu (1997)
• Children (?)	Chiswick and Repetto (2001), Dustmann (1994)
B. *Efficiency* (+)	
• Age at migration (-)	Chiswick and Miller (1995)(2007a), Espenshade and Fu (1997), Grenier (1984), Isphording and Otten (2011)(2012)
• Education (+)	Chiswick and Miller (1995)(2007a), Dustmann (1994), Espenshade and Fu (1997), Isphording and Otten (2011)(2012)
• Origin country language skills (+)	Dustmann (1994), Isphording and Otten (2011)
• Linguistic Distance (-)	Chiswick and Miller (2005a), Isphording and Otten (2011)(2012)
• Economic motive for migration (+)	Chiswick and Miller (2007a)(2008b)
C. *Economic incentives* (+)	
• Expected duration of stay (+)	Chiswick and Miller (2006)(2008b), Isphording and Otten (2011)
• Expected gain in earnings (+)	Dustmann (1994)

The evidence in Table 3 is important for the development of policy on immigrant selection and settlement. This will be discussed in Section 6. But it is also important in terms of

enhancing the understanding of many real world phenomena. Consider, for example, the very low level of English language proficiency among Mexican immigrants in the United States. The Mexican immigrants:

(1) have a very low level of schooling (an average of about eight years for adult men);
(2) have a high propensity to be sojourners, with substantial to and fro migration and often a short expected duration of stay;
(3) have low costs of migration because they come from an origin adjacent to the United States;
(4) are relatively recent immigrants as compared to the Europeans immigrants;
(5) tend to live in large Hispanic enclaves where they can live and even work in a Spanish language environment; and
(6) finally, are not skill tested for an immigration visa as they tend to be in the US under a visa for family reunification, under a formal or an informal amnesty for former undocumented migrants, or are in an illegal or unauthorized status.

These characteristics of immigrants from Mexico are all associated with lower destination language proficiency among immigrants in general.

5. Effects of Language on Earnings

The analyses of the consequences for immigrants of destination language proficiency have focused on labor market earnings (Carliner, 1981; Chiswick, 1991, 1998; Chiswick and

Miller, 2001, 2005b, 2007a; Dustmann, 1994; Dustmann and van Soest, 2001; Grenier, 1987; Kossoudji, 1988; McManus et al.,1983; Tainer, 1988)[16]. This focus has arisen for two reasons. One is the interest in earnings per se, as it is a key determinant of economic status and poverty. The other is the general availability of data on earnings in censuses and surveys that include information on immigrants' destination language proficiency, but not for many other outcome measures.[17]

5.1 Background Considerations

The analyses of earnings are generally performed primarily for adult (but non-aged) men because of the technical difficulties in estimating earnings equations for groups, such as women and aged men, who have relatively low labor force participation rates. The equations are usually estimated using the 'human capital earnings function', where the natural logarithm of earnings is regressed on a set of explanatory variables, typically including years of schooling, years of labor market experience (and its square) and variables for marital status and racial/ethnic origin. In analyses for the study of the impact of immigrant language proficiency, additional variables include duration in the destination, destination language proficiency and, sometimes, residence in

[16] While it would be desirable to have data on the language used in the immigrant's workplace, these data are generally not available. Moreover, immigrants may choose (or be chosen for) jobs that match their language skills, rather than the workplace causing language proficiency. One exception is the 2001 and 2006 Census of Canada data sets. These contain information on: (i) knowledge of the official languages (English and French) of Canada; (ii) other languages spoken; (iii) the language spoken most often at home; (iii) mother tongue (the language first learned at home in childhood that is still understood); (iv) the language used most often in the worker's job, as well as any other language used on a regular basis in the job. See Christofides and Swidinsky (2010) for an analysis of the links between language use and earnings among the native born in Canada based on these data obtained from the 2001 Census.

[17] For an analysis of the effects of dominant language proficiency (in this case Spanish) on the earnings of indigenous people in Bolivia, see Chiswick et al. (2000).

a linguistic concentration (enclave) area (Chiswick and Miller, 1995, 2005b).

Various destination language proficiency variables have been considered for inclusion in such a specification. These usually follow the way the data are collected in the majority of the data sets that are useful for study of labor market outcomes. The standard 'census' type question used in the US and Australia is outlined in Box 6.

Based on Box 6, it is apparent that self-reported English proficiency could be categorized using a five-interval scale. It is assumed that the highest level of proficiency on this scale would be 5 = Speaks only English at home, although there is no information on the proficiency of those who speak only English at home. All other levels relate to individuals who speak a language other than English at home, and self-report speaking English: 4 = Very Well; 3 = Well; 2 = Not Well; 1 = Not at All. Evans (1987), who converted the categorical information to a 'continuous' measure, proposes the score: 0 = Speaks no English at all; 33 = Speaks English 'Not Well'; 67 = Speaks English 'Well'; 100 = Speaks English 'Very Well' or speaks only English at home. This is based on Evans's findings that the effects of English proficiency variables on occupational attainment were approximately linear, with little difference in earnings between the 'English only' and the 'English very well' speakers. The combining of the two most fluent categories has support in the literature, see, for example, Kominski (1989), Espenshade and Fu (1997) and Bleakley and Chin (2004). Thus, Espenshade and Fu (1997, p.293) argue that '…there is not much difference in English proficiency between immigrants who use a language other than English at home but who say they speak English "Very Well" and those who use only English at home'.

Box 6: Typical Census language questions

11.**a** Does this person speak a language other than English at home?
□ Yes
□ No→ *Skip to 12*

b What is this language?

(For example: Korean, Italian, Spanish, Vietnamese)

c How well does this person speak English?
□ Very well
□ Well
□ Not well
□ Not at all

Source: U.S. Department of Commerce, Bureau of the Census, Census 2000 Long Form Questionnaire.

It can also be argued that as the information on dominant language skills is self-reported, and hence possibly subject to reporting errors, the categories specified in the census questionnaire should be further grouped. Hence, in many analyses a dichotomous variable for dominant language proficiency is defined, where, for example, workers who speak English only, or if a language other than English is spoken at home report speaking English 'very well' or 'well', are distinguished from other workers, with the first group being viewed as proficient in English and the latter group ('not well' and 'not at all') viewed as being deficient. Within this framework, some researchers move the group who speak only 'well' into the deficient category.

There is a difficult methodological consideration in the estimation of the dominant language-augmented human capital earnings equation outlined above. As discussed in Section 4, one of the influences on dominant language proficiency is the increases in earnings expected to be associated with improvement in language skills. In other words, in the study of earnings,

earnings are held to depend on the immigrant's proficiency in the dominant language. Yet proficiency itself is determined (in part) by the expected earnings payoff from becoming more proficient. Thus, earnings and proficiency are jointly determined.

Related to this, Dustmann and van Soest (2001) draw attention to the potential role of unobserved heterogeneity, where a factor which is important to both the development of dominant language skills and earnings determination is not observed by the researcher. The classic example is ability. This unobserved heterogeneity could be associated with either upward or downward bias in the OLS estimate of the effect of language skills on earnings (Dustmann and van Soest, 2001).

A further problem that has consequences for the correlation of the language variable and the error term in the earnings equation is the possibility that the self-reported language proficiency data suffer from misclassification errors, a factor that has motivated the use of binary indicators of dominant language proficiency in preference to the use of the more detailed information that is often available. The use of dichotomous indicators of proficiency may reduce the gravity of the errors in variables problem, but it will not eliminate it entirely. The misclassification errors that arise with self-reported language data could be either purely random, or they could persist over time. Time persistent errors will arise where each individual has an inherent tendency to consistently over-report or under-report their language ability.[18] Random misclassification errors will be associated with a bias toward zero in the estimated impact of language proficiency on earnings. Empirically, unobserved heterogeneity was shown by

[18] Berman, Lang and Siniver (2003) note that the relative importance of time persistent errors means that the language variable in their differenced model may have a lower noise-to-signal ratio than the standard language variable in a cross-sectional equation: in their study for Israel, these authors collected at the same time the information on current Hebrew proficiency and proficiency at the commencement of the current job.

Dustmann and van Soest (2001) to be associated with an upward bias, which was approximately offset by the negative bias due to misclassification errors. General measurement errors, which were modeled through examination of the correlation of the disturbance terms in the language proficiency and earnings models, were associated with a pronounced negative bias. Dustmann and van Soest's (2001) research suggests that the OLS estimates should thus be considered as lower bounds of the true effects of language skills on earnings.

This finding from the detailed study by Dustmann and van Soest (2001) has been echoed by various studies that adopted a conventional instrumental variables approach to the endogeneity problem. Under this approach a predicted language proficiency variable, rather than the observed proficiency, is entered into the earnings equation. When the dominant language-augmented human capital earnings equation has been estimated using an IV approach, the coefficient on the instrumented language variable is extremely large, far too large to be believed. See Box 7. This econometric problem arises when the instrumented (predicted) variable is dichotomous and the residuals in the auxiliary equation and the main equation are positively correlated because some of the same omitted variables (for example, ability or childhood home environment) are in both equations. The problem of disentangling the endogeneity of earnings and language proficiency has not been resolved (Chiswick and Miller, 1995; Dustmann and van Soest, 2001).

Box 7: Results Using Instrumental Variables

A number of researchers have attempted to accommodate the endogeneity between earnings and destination language proficiency using an instrumental variables (IV) approach. Shields and Wheatley Price (2002, p.158) offer a concise summary of the results, which we repeat in full here. Please refer to the original for the relevant references.

> Chiswick and Miller (1992), for example, found an increase in the partial effect of language fluency on earnings (*t*-ratios in parentheses) from 0.169 (12.52, OLS) to 0.571 (5.43, IV) using 1980 United States data and veteran status, foreign marriage, children and minority language concentration measures as identifying instruments. They also noted a change from 0.122 (2.43, OLS) to 0.414 (1.34, IV) amongst immigrants in 1981 Canadian data with foreign marriage and minority language concentration measures as identifying instruments. In Australia the results changed from 0.052 (2.52, OLS) in 1981 and 0.083 (4.75, OLS) in 1986 to -0.243 (1.20, IV) and 0.043 (0.52, IV), respectively, with foreign marriage, number and age of children and minority language concentration as identifying instruments. Chiswick (1998) found an increase from 0.111 (12.66, OLS) to 0.351 (4.25, IV) using 1983 data from Israel using Tel Aviv, Jerusalem, foreign marriage, number of children and minority language concentration measures as identifying instruments. Finally, Dustmann and van Soest (1998a), using German socio-economic panel data between 1994-1993 found a language effect on earnings increase from 0.0538 (7.08, OLS) to 0.155 (2.28, IV) with father's education measures as identifying instruments.

Using UK data, Shields and Wheatley Price (2002) reported increases in the effects of fluency on earnings from 0.0887 (3.62, OLS) to 0.1651 (3.96, IV) using the languages in which the interview was conducted as instruments, and from 0.0887 (3.62, OLS) to 0.1142 (0.99, IV) using married to a UK born spouse and the number of dependent children in the household as instruments.

With these considerations in mind, we turn to the empirical evidence. The volume of research precludes an exhaustive review. Rather, we travel the globe, examining a selection of studies from different destination and origin countries (and languages) as well as different time periods, along with different measures of language proficiency. Despite these differences, there seems to be a striking common finding in the mainstream studies: dominant language proficiency among immigrants is rewarded rather handsomely around the globe.

5.2 Around the Globe: Different Countries, Different Languages, Similar Results

Australia

The study of the determinants and consequences of the English language proficiency of migrants is of particular interest for Australia. In various formats, throughout the 20th century, English language tests have played an important role in the admission of immigrants. Under the Skilled Independent Visa category, points are currently awarded on the basis of proficiency in English, among other personal characteristics.[19] A recent increase in the English language requirements for the skill-based independent immigrants, but not for other applicants, provided an opportunity to demonstrate the effect of immigration policy on the skills of immigrants (Chiswick and Miller, 2006). The English proficiency of the skill tested immigrants increased with no change for the others.

The early research on the links between English language skills and labor market outcomes in Australia was based largely on census data, with the 1981 Census marking the first release of unit record Census data. Later research has been based largely on the Longitudinal Survey of Immigrants to Australia.

Two of the early studies for Australia based on census data were by Evans (1987) and Chiswick and Miller (1985). Both used data from the 1981 Census. In Evans' research the focus was on the determinants, including language skills, of occupational status among male immigrants from Mediterranean, Northwest European, Third World and Eastern European countries. The English language proficiency information was collected using a format the same as Box 6, though as tests showed that the effects of the various levels of proficiency were

[19] See Appendix 1 of Kan (1991) for information on changes in Australia's immigration policy leading up to 1989. As noted by Kan (1991), numerical scoring was used as an administrative arrangement as early as 1979. The importance of the change in 1989 was that the points system was recognized in law.

approximately linear, a continuous measure was used, ranging from 0 where the immigrant did not speak English, to 100 where the immigrant either spoke English only or, where a language other than English was spoken at home, English was spoken very well. The dependent variable was an occupational status score. In each instance better language skills were associated with statistically significant increases in occupational status scores. The increases in occupational status associated with improved English language proficiency differed across the four birthplace groups, however, being stronger for those from Third World and Eastern European countries than they were for Mediterranean and Northwest European countries. Evans (1987) noted that Third World and Eastern European countries were, at the time of data collection, not associated with immigrant enclaves in Australia, and thus argued that this pattern across birthplace regions showed that enclave economies can shelter immigrants with poor language skills from adverse labor market consequences. We return to this theme in Section 5.4.

The study by Chiswick and Miller (1985) was more conventional in terms of its focus on earnings, and its use of dichotomous variables for whether the immigrant spoke a language other than English at home, and for whether an immigrant who spoke a language other than English at home spoke English either not well or not at all. Separate analyses were undertaken for male immigrants from English-speaking countries and for male immigrants from non-English-speaking countries. It was reported that among immigrants from English-speaking countries, language skills did not influence earnings, most likely because nearly all were proficient. However, among immigrants from non-English-speaking countries, those who spoke a language other than English at home but spoke English very well or well had earnings 4.8 percent lower than monolingual English speakers, while those with poor English skills had a further 6.7 percent

earnings disadvantage. In other words, there was an earnings differential of around 12 percent between those with poor English skills and monolingual English speakers.

In the 1990s and 2000s the Australian government department responsible for immigration undertook three longitudinal surveys of immigrants. These, respectively, followed immigrants who arrived in Australia between September 1993 and August 1995, arrived between September 1999 and August 2000, and either arrived in Australia, or were granted a permanent visa in Australia, between December 2004 and March 2005. The surveys collected very detailed data on the respondents' immigration and settlement experiences, including data on language skills, and they have provided the basis for numerous studies, including those in the research volume edited by Cobb-Clark and Khoo (2006). Chiswick, Lee and Miller (2005c) use data from the first of these surveys to study the impact of language skills on recent immigrants' earnings. The three waves of this data collection, conducted six months, 18 months and 42 months after arrival, were used in the analysis. The English Language Proficiency variable was dichotomous, having a value of one for those who speak only English, speak English best or speak English very well. Various sets of estimates were presented (OLS, selection bias corrected, inertia panel data models), though the set that is most comparable with the literature is the OLS estimations undertaken for each wave of data collection. These results revealed that proficiency in English was associated with between 19 and 24 percent higher earnings.[20]

Thus, the research for Australia, using census and other data sets, and covering several decades, consistently reports that proficiency in English is associated with a considerable

[20] These results are of further interest as the estimating equations employed in this study took account of visa class, a variable that is generally not available in the Censuses and surveys used to study immigrant labor market adjustment.

earnings premium, of at least 10 percent when all immigrants are considered, and possibly as high as 20 percent when the focus is upon recent immigrants.

Canada

Canada is a bilingual country, with English and French being the official languages, though French is predominately spoken in Quebec and English is spoken by the overwhelming majority in the rest of Canada. This official bilingual status, and the tensions between French-speaking Quebec and English-speaking 'Rest of Canada', have provided the background for the development of a rich body of literature examining language issues. Much of this has focused on the native born, though a somewhat separate literature on immigrants has developed, written mainly by scholars interested in immigrant adjustment. Reflecting the availability of data, the basis of this research in Canada has been similar to that in Australia, with the early research relying on census data, and recent research using a number of more specialized data sets. The analyses for Canada are of special interest due to the findings that emerge in a dual-language labor market, and when some of the specialized data sets are used.

Census of Canada data sets have considerable depth of information on language use and proficiency (see footnote 16), and the studies using these data have differed appreciably in the way the language variables are constructed. Most studies have focused on either the official language information, or on this information in conjunction with home language use or region of residence. The rationale for this approach is described by Carliner (1981, p.388) as: 'Since the purpose of this paper is to analyze wage premiums for language skills rather than shifts in language use, it seemed preferable to classify workers by language currently used rather than by language first used'.

Carliner's (1981) analyses, based on 1971 census data, covered observations pooled across the native born and foreign born, and incorporated intercept shift terms for a number of immigrant arrival cohorts. The census information on language spoken at home (which Carliner (1981) terms mother tongue for convenience of expression) was combined with that on ability to conduct a conversation in English and French to construct eight language groups, such as monolingual English speaker, native English speaker who also speaks French, and monolingual speakers of other languages who spoke neither English nor French. The latter group who could speak neither English nor French had the lowest earnings (compared to being able to speak an official language) in both Quebec and the rest of Canada, though the disadvantage was three-times larger in Quebec (coefficient of -0.281) than in English Canada (coefficient of -0.107). The increments in earnings associated with the acquisition of official language skills differed between French and English, and for Quebec and English Canada. For example, in Quebec the earnings differential between the group who could speak neither English nor French and non-native French speakers was 10 percentage points greater than that between those without official language skills and non-native English speakers. In English Canada, the difference in the earnings outcomes for these comparisons was only three percentage points. It is not clear, however, whether these patterns would apply if the foreign-born component of the sample was analyzed separately.

This issue is addressed in Chiswick and Miller (1988). They interact the official language knowledge information with region of residence (Quebec versus the rest of Canada), using data from the 1981 Census of Canada. Separate analyses were conducted for the foreign born. The results showed that the foreign born who could not speak English or French had earnings almost

20 percent below the earnings of English-French bilinguals.[21] Monolingual English speakers earned less than English-French bilinguals in both Quebec and English Canada, whereas monolingual French speakers earned less than English-French bilinguals only in Quebec.

A more recent census data-based study which has a structure similar to Chiswick and Miller (1988) is by Nadeau and Seckin (2010). Thus their study of adult male immigrants was based around earnings equations estimated separately for Quebec and the rest of Canada. They used data from the 1981, 1991 and 2001 censuses. These data sets were examined separately. Information on both the language spoken at home and on the ability to conduct a conversation in an official language was incorporated into the estimating equation, with separate sets of variables used to record each of these language attributes. As around 40 percent of immigrants in 2000 spoke a language other than English or French at home, and only around two percent could not speak either English or French, it is apparent that these variables will contain different information.

Within the rest of Canada, those who could not speak either English or French earned around 12 percent less than monolingual English speakers in 2000, and 17 percent less than bilinguals. This disadvantage had increased slightly since 1980. There was a further disadvantage associated with speaking neither English nor French at home, of 11 percent, which also was greater than the comparable disadvantage in 1980. It can be inferred that the disadvantage in the labor market associated with a lack of official (English or French) language skills is over 20 percent in English Canada.

In Quebec, the only official language category that was distinguished on the basis of

[21] The variable for the absence of official language skills was not interacted with region of residence.

earnings among the foreign born adult males was the English-French bilinguals, who had earnings around 12 percent more than the other official language groups in 2000. However, the use of both English and French in the home, along with the use of a language other than English or French, was associated with significantly lower earnings in 2000. The estimated effects, however, did not exhibit any pattern, in terms of magnitude or statistical significance, across the three data sets analyzed. Nevertheless, the point estimates suggest that those who could not speak English or French, and hence used neither of these languages at home, have earnings 6-14 percent less than English monolinguals in Quebec.

Turning to studies that use specialized data sets, Ferrer, Green and Riddell (2006) focused on the role of objective measures of document and quantitative literacy, using the Ontario Immigrant Literacy Survey of 1998 and the Canadian 1994 International Adult Literacy survey. Document literacy refers to the individual's capabilities in locating and using information. Quantitative literacy refers to problem solving in real world contexts. Given the collinearity between these measures, the emphasis in the study was on an average of these variables. In addition, the authors use a self-reported measure of language skills, and construct a dichotomous variable that is set equal to one where the individual could only express himself poorly or not at all in English or French. This subjective measure of language skills was associated with around a 30 percent earnings disadvantage, but this earnings disadvantage was reduced by around one-third when the objective measure of literacy was included in the estimating equation.

The objective measure of literacy, given by the average of the document and quantitative scores, was associated with significantly higher earnings. The effect for immigrants educated

abroad was approximately the same as that for the native born, and the effect for immigrants educated in Canada was higher than that for the native born. A 100 point increase in the literacy score (which corresponds to a 1.5 standard deviation increase in the score that is measured on a 0-500 scale) was associated with 30 percent higher earnings among immigrants educated abroad, and 50 percent higher earnings among immigrants educated in Canada.

The research for Canada thus shows that proficiency in English or French is associated with an earnings premium of at least 20 percent, and perhaps as high as 30 percent. English-French bilingualism is not necessarily associated with economic rewards in these studies, even in this dual language country.

Germany

Germany, as with Israel, has a dominant language which is not an international language. For this reason it is of considerable interest to examine the links between knowledge of German and labor market outcomes among immigrants in that country.

Dustmann (1994) uses data from the 1984 German Socio-Economic Panel. Interesting features of this study are the statistical controls for both writing and speaking proficiency as separate variables. In each case good and very good skills were distinguished from intermediate skills, and from the benchmark group of those who spoke (or wrote as the case may be) German badly or not at all. Among males, migrants with good or very good speaking skills were associated with almost 7 percent higher earnings than those with poor German language skills, and those with intermediate speaking skills had an earnings advantage of around one-half of that amount. Among females, only those in the highest speaking proficiency category were

distinguished by having significantly higher earnings. Their earnings advantage (7.1 percent) over migrants with limited or no German speaking skills was similar to that for males.

Writing abilities were also associated with statistically significant earnings effects among migrants in Germany. However, only those with good or very good writing skills were characterized by statistically significant earnings effects, of 7 percent in the case of males, and 15 percent in the case of females.

Dustmann (1994) also included variables for both writing and speaking skills in an encompassing model. For males the result was slightly smaller estimated earnings effects associated with each skill. Among females, the effect was that the writing skill variable was statistically significant whereas the speaking skill variable was not. These variables, however, will be highly correlated. Dustmann (1994, p.151) notes 'Since a migrant who is fluent or very fluent in German writing should also be quite fluent in spoken German…'

While Dustmann (1994) used only the 1984 wave of the German Socio-Economic Panel, multiple waves of this data collection over 1984-1993 were employed by Dustmann and van Soest (2001)(2002). Both studies have a focus on addressing endogeneity bias, and use only the information on speaking fluency. The 2001 study examines the impact of German speaking skills on earnings for males only, whereas the 2002 study contains separate sets of estimations for males and females. The studies also differ in their approaches. In Dustmann and van Soest (2001) the five categories in the speaking skills data were collapsed to three: (i) good and very good; (ii) intermediate; and (iii) bad and very bad, to ensure each group was numerically important. This follows Dustmann (1994). The probability of being in these German speaking skills groups was examined using an ordered probit model, and the latent index from this

probability model was used as the language variable in the earnings equation. In Dustmann and van Soest (2002), only a dichotomous German speaking skills variable was used, where those with good or very good German speaking skills were distinguished from the other proficiency categories. This approach was taken to provide a specification that was similar to that used in studies for other countries.

The patterns of results in the two studies of Germany are similar, and can be described using the more conventional 2002 study. In a standard earnings equation estimated using OLS, German speaking fluency was associated with about five percent higher earnings for males.[22] Taking account of unobserved heterogeneity resulted in a slightly lower estimate of the earnings premium to speaking skills. Taking account of time-varying measurement errors was associated with a pronounced increase in the estimated effects of speaking proficiency on earnings, to around 14 percent in one model. However, time-persistent measurement errors did not appear to be overly important.[23]

Hence, the research for Germany indicates an earnings premium to German speaking proficiency of at least five percent. The more rigorous research undertaken by Dustmann and van Soest (2001)(2002) to address the endogeneity issue suggests that the earnings premium could be three-times higher than this, at around 15 percent.

[22] Dustmann and van Soest (2002, p.484) note that the OLS estimate of a five percent increase in earnings associated with dominant language proficiency is lower than that reported in other countries, and suggest that this could be associated with their inclusion of the 'intermediate' speaking skills group in with those who have poor speaking skills. The evidence reported in Dustmann (1994), based on a more general specification that contained variables for (i) good and very good; (ii) intermediate; and (iii) bad and very bad German speaking skills, indicates that the Dustmann-van Soest procedure would tend to decrease the estimated effect of dominant language proficiency.

[23] The results reported for women were similar to those discussed for men.

Israel

Analysis of the role of language skills in Israel is of interest for three main reasons. First, unlike English which is the dominant language in the majority of the immigrant-receiving countries covered in the literature, although Hebrew is the language of religious practice among Jews worldwide, it is not otherwise an international language. Second, Israel is not as developed as countries such as Australia, Canada, Germany, the United Kingdom and the United States that are typically studied in this field of research. Third, the more recent studies of immigrants in the Israeli labor market cover the period following the large inflow of immigrants from the Former Soviet Union (FSU).[24] Very few arrivals knew any Hebrew because the study of Hebrew had been prohibited in the Soviet Union.

Chiswick and Repetto (2001) examine the returns to Hebrew writing and speaking skills using data from the 20 percent microdata file of the 1972 Census of Israel.[25] They have a focus on foreign-born adult Jewish males. The main language variables were a dichotomous variable for the ability to write a simple letter in Hebrew, and a four-category speaking variable that distinguished among (i) Hebrew only; (ii) Hebrew was the primary language and the individual speaks other languages; (iii) Hebrew was a second language; and (iv) does not speak Hebrew. The results revealed that the earnings of those who speak only Hebrew and those bilinguals for whom Hebrew was the primary language were not significantly different. Compared to these men, however, earnings were eight percent lower among bilinguals who reported Hebrew as

[24] In 1989 there was a policy shift in the Soviet Union, and restrictions on the movement of Jews to Israel were removed, and rapidly resulted in a more than 15 percent increase in the Israeli population and labor force.

[25] The 1972 Census of Israel had the richest set of questions on language proficiency in the Israeli Censuses. The 1983 Census had only one question, and the subsequent Censuses had none. Analysis of the 1983 Census question on language provided similar findings as the 1972 Census analysis (Chiswick, 1998; Chiswick and Repetto, 2001).

their second language, and 21 percent lower among those who could not speak Hebrew.[26]

Similarly, workers who could write a simple letter in Hebrew had earnings around 12 percent higher than those who could not perform this task. An interesting feature of the results was that when both Hebrew writing and speaking skills were entered into the estimating equation at the same time, both were statistically significant: Hebrew writing skills were associated with around nine percent higher earnings, and Hebrew speaking skills with around 11 percent higher earnings. In other words, immigrants who could both speak and write in Hebrew had earnings around 20 percent higher than those who did not possess these skills.

Berman, Lang and Siniver (2003) used a special 1994 survey to examine the earnings among male FSU immigrants in Israel. Their study therefore covers a much later period than the research by Chiswick and Repetto (2001). The survey contained information on Hebrew speaking skills on a five-point scale (not at all, a little bit, not so well, well, and very well), which the authors used as a continuous, cardinal measure. Information was collected (simultaneously) on wages and language skills at the time of interview and at the time they started their current job. Also, data were collected for workers in both high-skilled and low-skilled occupations.

Four main findings were reported. First, the continuous variable recording proficiency in Hebrew indicated a 26 percent difference in wages between the extremes of the proficiency scale (not at all and very well). Second, when separate dichotomous variables were used for each of

[26] Other variables being the same, among adult male immigrants in Israel, having been born in an English-speaking developed country is associated with lower proficiency in Hebrew and with higher earnings. Even after controlling for country of birth, speaking English on a daily basis is associated with a highly statistically significant 15 percent earnings advantage (Chiswick and Repetto, 2001). Further research is needed to determine whether this reflects the role of English as an international language, the high earnings potential of English speakers in the advanced English-speaking countries, or some other factor.

the language proficiency groups, there was a clear hierarchy of earnings effects in comparison to the reference group of not being able to speak Hebrew, from 31.4 percent in the top, very well, proficiency category, 19.5 percent for the second highest, well, category, 14.8 percent for the not so well category, and 8.6 percent for the second lowest category of being able to speak Hebrew a little bit. Hence, while these earnings increments are not quite on a linear scale, the linear scale provides a reasonable approximation. Third, this wage effect was diminished only slightly in a differenced model that allowed for ability bias. Fourth, the payoff to Hebrew fluency was higher in the skilled occupations than in the unskilled occupations: indeed proficiency in Hebrew had little effect in unskilled occupations. This suggests a complementarity between language and occupational skill in the generation of earnings (Chiswick and Miller, 2003)

The findings of Chiswick and Repetto (2001) and Berman et al. (2003) therefore indicate an earnings gap of at least 20 percent between workers who are proficient in Hebrew and those who lack this skill. There is also evidence of a complementarity between Hebrew language skills and the other human capital needed for entry into skilled occupations. It is noted, however, that these results are based on data collected in 1972 and 1994, reflecting the surprising absence of information on Hebrew language skills in more recent censuses. Further comment on data collection issues is provided in Section 6.

Spain

Research into the effects of dominant language proficiency among immigrants in Spain can offer evidence that complements studies for other countries. This is because of the different dominant language (Spanish), and the fact that the large flows of immigrants into Spain are of

relatively recent origin. Budriá and Swedberg (2012) note that the number of foreign workers present in Spain increased by two and one-quarter million during 2001-2008. However, until the Spanish National Immigrant Survey undertaken during November 2006 to February 2007, there were no nation-wide data that would facilitate study of the effects of proficiency in Spanish on the earnings of immigrants. Budriá and Swedberg (2012) use these data in OLS and IV estimations. The data on proficiency in Spanish were derived from a question which asked respondents to rate their speaking proficiency in the context of the skills needed for communicating at work, at the bank, and with the public authorities/administration. Four response categories were provided, ranging from 'very well' to 'need to improve'. Budriá and Swedberg (2012) used these data to construct a dichotomous Spanish language proficiency variable, where immigrants who could speak Spanish very well were distinguished from other immigrants. The Spanish proficiency rate was around two-thirds.

The OLS estimates indicated that adult male immigrants who were proficient in Spanish had 4.8 percent higher earnings than those who lacked this skill.[27] This language premium is much lower than those generally reported in studies for other countries. It is noted, however, that the partial effect of schooling on earnings was only 1.1 percent, which is also much lower than the return typically found for other countries. This suggests very low payoffs to skill in the Spanish labor market

The IV estimates were identified using information on age at arrival in Spain, the presence of a child who is proficient in Spanish, and by plans to stay in Spain for the next five years. Proficiency in Spanish was typically associated with around 25 percent higher earnings in

[27] One aspect of the estimating equation that needs to be noted is that account was not taken of duration of the immigrant's stay in Spain. The inclusion of a years since migration variable in the model was reported to result in little change to the findings.

the IV estimations. This marked change between the earnings premium to dominant language proficiency obtained under OLS and IV estimations is similar to that found for other countries: see Box 7.

In IV estimations undertaken for separate samples of better educated and less-well educated workers, proficiency in Spanish was associated with almost 50 percent higher earnings among the better educated, whereas it was associated with around 20 percent higher earnings among the less-well educated. The OLS estimates indicated six percent higher earnings among the better educated and a statistically insignificant three percent higher earnings among the less-well educated. This evidence supports the complementary of human capital skills reported by Chiswick and Miller (2003).

It is apparent from this overview of the first nationwide study of the impact of dominant language proficiency and schooling on the earnings of immigrants in Spain that the returns to human capital in Spain are very low. The large difference between the OLS and IV estimates for the effects of Spanish proficiency is similar to that reported in studies for other countries, and suggests that the OLS findings could be a considerable under-estimate of the true impact. Further research, using alternative datasets, is needed on this topic.

United Kingdom

In 2000, around 9 percent of the working age population of Britain was born overseas (Dustmann and Fabbri, 2003). Many of these immigrants are classified as belonging to racial and ethnic minorities (Black Caribbeans, African Asians, Indians, Pakistanis, Bangladeshis and Chinese). Despite this importance, there has been comparatively little research on the effects of English language proficiency on earnings in the UK. As Dustmann and Fabbri (2003, p.697)

explain in their fairly recent study, 'The data sources we use for this analysis are to our knowledge the only data sets for the UK that contain information about immigrants' language proficiency, as well as information on employment status and earnings'. The often cited earlier study by Shields and Wheatley Price (2002) was based on an occupational status score, specifically the mean wage of the occupation of employment. Shields and Wheatley Price (2002) captured only the across-occupation gains in earnings associated with proficiency in English and not the within-occupation gains. They find that English proficiency was associated with around nine percent higher earnings.

Dustmann and Fabbri (2003) examine data from two surveys, which were collected in the early-to-mid 1990s. One of these surveys contained self-assessed information on speaking, reading and writing skills, and the other contained interviewer's evaluations of the respondents' spoken language proficiency. Both employment and earnings outcomes were examined. The authors accommodate the bias associated with unobserved heterogeneity using a matching estimator, and the bias associated with measurement error using an IV type approach. Binary indicators of English proficiency were used. Workers with an English mother tongue, and those with a non-English mother tongue who reported speaking English good or very good were defined as proficient when the self-assessed data were used, and those assessed by the interviewer as being fairly fluent or speaking English fluently were defined as proficient under the alternative measure.

They report that dominant language speaking skills were significantly and positively related to the employment probability in an OLS estimation, and writing skills were also related to this labor market outcome, though the partial effects were stronger. When both sets of skills

were included in a single equation writing skills were statistically significant whereas speaking skills were not. The estimated effects associated with the language proficiency variables were smaller under the matching approach, though they remained statistically significant, whereas they were larger, though statistically insignificant under the IV approach. In general, being proficient in English was associated with an increase in the employment probability of between 10 and 20 percentage points.

The estimation of a standard human capital earnings equation using OLS showed that speaking proficiency was associated with around 20 percent higher earnings, and writing proficiency with around 15 percent higher earnings. Taking account of unobserved heterogeneity resulted in a higher (significant) estimate in one sample, and a smaller (insignificant) estimate in the other sample. The estimated effect on earnings of proficiency in English was much higher under the IV type approach (partial effect of 0.356 for the total sample, 0.460 for males and 0.844 for females), though each of these estimates was statistically insignificant.

Thus, according to the Dustmann-Fabbri (2003) study, dominant language proficiency among immigrants in the UK is associated with substantial improvements in labor market outcomes. The standard OLS estimates of these, of 10-20 percentage points improvement in the employment probability, and up to 20 percent increase in earnings, are comparable to findings in other highly developed countries.

Miranda and Zhu (2013) based their analyses on the UK Household Longitudinal Survey 2009-2011, and focused on adult males. Their language variable is a measure of deficiency in English, and is termed ‘English as Additional Language’. This was defined from the responses to the question ‘Is English your first language?’. There was a 16 percent wage differential between

those with good and deficient English when estimated using OLS. When IV was used, with language of the country of origin and interactions of this with age-at-arrival as instruments (see Bleakley and Chin, 2004) the estimated wage differential increased to 25 percent. This estimate was robust to the tests undertaken, which involved the choice of instruments (relying only upon the interaction terms with age-at-arrival) and sample (immigrants only rather than a pooled sample of immigrants and the native born).

Thus, while there has been limited research in the UK compared to other major immigrant receiving countries, the studies available reveal that English language proficiency is associated with 15-20 percent higher earnings. Miranda and Zhu's (2013) IV estimates, which appear to be robust, indicate that the earnings differential could be as high as 25 percent.

United States

Immigrant flows into the US exceed those into any other country. During the 2000s, for example, more than one million people achieved immigrant status each year. Much of the research on the labor market rewards for dominant language proficiency has been undertaken on immigrants to the US. Initially, this research was concentrated on understanding labor market outcomes among Hispanics. However, as the mix of immigrant arrivals became more diversified following the 1965 Amendments to the Immigration and Nationality Act that removed the severe restrictions on Asian immigration, the coverage of the research broadened. Among what may be termed the milestone studies, the early research was based on the 1976 Survey of Income and Education, whereas the later research has made use of census data.

McManus, Gould, and Welch (1983) use the 1976 Survey of Income and Education in an analysis of the earnings of Hispanic men. They attempt to use much of the information on

language use and proficiency in the eight separate questions in the survey. They discard many of the variables as redundant and create an index of English language deficiency. Deficiency in English was shown to have negative effects on earnings, where the negative effects increased with higher levels of schooling and experience. When their index of English language deficiency was included in the equation, other variables for immigrant characteristics became statistically insignificant, most likely because of problems created by the construction of the language index.

Kossoudji (1988) also based her analysis on the 1976 Survey of Income and Education. The focus was on foreign-born men of Hispanic and East Asian origin. Unlike McManus, Gould and Welch (1983), however, and perhaps setting the standard for subsequent analyses, Kossoudji (1988) adopted a simple categorization of the language information, defining only three English proficiency groups: (i) fluent English speakers; (ii) those able to communicate in English; and (iii) those who speak little or no English. Kossoudji looks at the effects of English proficiency on occupational states and on earnings within broad occupational levels. She shows that those with the poorest English skills tend to be employed in lower-level jobs, and that this decrease in occupational status is more pronounced for Hispanics than it is for Asians. Differences in occupational attainment accounted for a negligible part of the overall earnings disadvantage of Asian immigrants with limited English skills, and around one-third of the disadvantage of Hispanic immigrants with limited English skills. English language skills were not statistically significant in the most skilled occupations. Within the low-skilled occupations, only those with the poorest skills tended to be disadvantaged, and this disadvantage was greater for Hispanics than for Asians. The within occupational wage effects of limited English proficiency at the lowest skill levels were around 20 percent for Hispanics, and 10 percent for

Asians.

Tainer (1988) also used the 1976 Survey of Income and Education for her analysis of the effect of English language proficiency on earnings. She found the effects to be positive and significant for foreign-born men, but there were variations in magnitude by ethnicity. In order to study these effects, Tainer created two language proficiency variables. The first variable, *SPEAK*, was based on the question, "How well do you speak English?" where proficiency was self-ranked on a scale of 1 to 5. The second variable, *INDEX*, was a continuous index (0 to 1) of three variables: proficiency (*SPEAK*), usage at home, and ability to understand English. Since these measures were not directly comparable, Tainer compared their elasticities and found that for foreign-born men, a 1 percent improvement in *SPEAK* increased annual earnings by 0.08 percent, and a 1 percent improvement in *INDEX* increased annual earnings by 0.04 percent (Tainer, 1988, p.117). Tainer also compared results for Europeans, Asians, and Hispanics in an effort to distinguish between the effects of language proficiency for different groups. However, for both proficiency variables, only the coefficients for Hispanics were statistically significant at a 5 percent level (coefficients of 0.17 for *SPEAK* and 0.69 for *INDEX*). Tainer concluded that the earnings of Hispanic and Asian men were more sensitive to English language proficiency than those of European men.

Among the many studies using census data are Chiswick and Miller (2002) and Bleakley and Chin (2004). Chiswick and Miller use 1990 census data in a detailed study of the earnings of male immigrants. Several specifications of the earnings equation were considered, including models with a single dichotomous variable recording fluency in English which was set equal to one if the immigrant speaks only English at home or speaks English 'very well' or 'well', and

with separate variables for each of the census English proficiency groups (see Box 6). The models also examined the role of linguistic enclaves, and these results are discussed in Section 5.4.

The results revealed that male immigrants who could not speak English, or who spoke it 'not well' had earnings around 19 percent lower than monolingual English speakers. Workers who spoke English only 'well' were at a nine percent earnings disadvantage compared to monolingual English speakers, whereas those who spoke English 'very well' were at a modest two percent earnings advantage compared with monolingual English speakers. When the single dichotomous variable was used to record proficiency in English, the results showed that workers who were proficient in English earned 14 percent more than those who were not proficient. This premium was of the same order of magnitude as was reported when a similar model was applied to the 1980 US census data (Chiswick and Miller, 1995). The increase in earnings associated with proficiency in English was shown to vary across countries of origin, and ranged from statistically insignificant effects for South Asia and Sub-Saharan Africa, to values of about 20 percent for immigrants from China and Japan. The increase in earnings associated with a greater proficiency in English comes about largely from shifting jobs to occupations requiring greater English language proficiency, with a minor role played by higher earnings in the same occupation (Chiswick and Miller, 2012).

A further indication of the heterogeneity of the earnings effect associated with proficiency in English in the 1990 census data is found in the study by Bleakley and Chin (2004). They focused on childhood immigrants, defined as those who were under the age of 18 at the time of arrival in the US. Their further restrictions, in order to get a more homogeneous group,

led to a focus on 25-38 year olds. Thus, the shortest duration in the US in their sample would have been seven years. They used a cardinal representation of the census English proficiency information (see Box 6), with the value rising with greater proficiency. Much of their analysis was based on models that did not control for educational attainment: in these the English proficiency variable had an extremely large effect on earnings. Once educational attainment was taken into account, however, the effects of English-speaking proficiency were much more modest, with one representative set of results indicating a difference in earnings of only around 6 percent between those who could not speak English at all and those who could speak English very well.[28] Bleakley and Chin (2004) address the problem of the potential endogeneity of English-speaking proficiency using an IV estimator, with age at arrival interacted with birthplace region as the identifying instrument. Consistent with results for the US and UK discussed above, the IV estimates were typically 50 percent, or more, greater than the OLS estimates.

Earnings effects of the magnitude reported in studies of data sets covering the 20th century also characterize the contemporary US labor market. This is revealed from analysis of data from the American Community Survey (2005-2009). Thus, Table 4 presents results from a regression analysis by nativity for the earnings of adult men. Among the foreign born, those who speak another language at home but who speak English 'very well' earn about 1 percent less than those who speak only English, while those who speak English only 'well' earn nearly 24 percent less. Earnings are even lower for immigrants who are even less proficient. These earnings differentials associated with English-speaking skills for 2005-2009 are much larger, by the order

[28] Bleakley and Chin (2004) argued that this showed that much of the effect of English-language proficiency on earnings is mediated through higher educational attainment, and this has specific policy implications. This is not as great a concern among immigrants who arrival as adults, which is around two-thirds of the foreign-born population in the US.

of 50 percent, than the earnings effects reported by Chiswick and Miller (2002), based on analyses of data for 1990. However, over the same time period the payoff to years of schooling for the foreign born increased from 4.3 percent to 6.0 percent, a 40 percent increase, and the return from schooling for the native born also increased. Hence, this change appears to be part of a pattern of a general increase in the effects on earnings of human capital among immigrants and the native born in the US.

Table 4: Analysis of earnings by nativity for adult males, United States, 2005-2009

Variable	Native Born	Foreign Born
English very well	-0.009	-0.013
	(2.94)	(4.23)
English well	-0.033	-0.236
	(5.23)	(66.24)
English not well/not at all	-0.003	-0.331
	(0.36)	(82.4)
Years of schooling	0.122	0.060
	(622.29)	(195.88)
Labor market experience	0.044	0.015
	(220.4)	(37.78)
Labor market experience squared/100	-0.075	-0.022
	(195.76)	(29.24)
Log weeks worked	1.145	1.017
	(993.26)	(384.38)
South	-0.046	-0.065
	(47.34)	(29.9)
Married (spouse present)	0.272	0.217
	(260.64)	(95.26)
Years since migration (YSM)	(a)	0.014
		(50.03)
YSM squared/100	(a)	-0.019
		(35.73)
Minority language concentration (fraction of state population)/100	-0.168	-0.484
	(8.99)	(38.18)
Constant	3.850	5.476
	(710.10)	453.99

Number of Observations	2763924	529773
R-squared	0.3969	0.3959

Notes: Dependent Variable = Natural logarithm of annual earnings; t-ratios in parentheses; (a) = Variables not entered.
Source: American Community Survey, United States, 2005-2009, microdata file.

5.3 *An Overview of the Effects of Dominant Language Proficiency on Earnings*

Table 5 presents a summary of the findings described above on the effects of dominant language proficiency on the earnings of immigrants.

Table 5: The Effects of Dominant Language Proficiency on the Earnings of Adult Male Immigrants[(a)]

Country	Typical increases in earnings associated with dominant language proficiency (%)	Illustrative studies
Australia	10-20	Chiswick and Miller (1985), Chiswick Lee and Miller (2005c), Evans (1987)
Canada	20-30	Carliner (1981), Chiswick and Miller (1988), Ferrer, Green and Riddell (2006)
Germany	5-15	Dustmann (1994), Dustmann and van Soest (2001)(2002)
Israel	10-25	Chiswick and Repetto (2001), Berman, Lang and Siniver (2003)
Spain	5	Budría and Swedberg (2012)
United Kingdom	15-20	Shields and Wheatley Price (2002), Dustmann and Fabbri (2003), Miranda and Zhu (2013)
United States	10-20	McManus, Gould and Welch (1983), Koussoudji (1988), Tainer (1988), Chiswick and Miller (2002, 2012)

(a) = Based on OLS analyses.

Clearly from the Table 5 evidence, the acquisition of dominant language proficiency estimated from OLS analyses is associated with higher earnings. But is investment in destination language proficiency profitable for immigrants? Considering only the labor market earnings impacts, a 15 percent increase in earnings per year from going from 'not proficient' to

'proficient' would imply a 30 percent rate of return on the investment if the cost of the language training involved the equivalent of a half of a year of full-time earnings, a 15 percent rate of return if it required a full year and a 7.5 percent rate of return if required two full years. Even if it required two full years, this is a high rate of return on the investment. Yet, this computation does not take into account the consumption, social and civic benefits, or the lowering of the costs of other investments in human capital. Thus, it appears that the investment in destination language proficiency is a profitable investment for immigrants and for society.

5.4 Ethnic Enclaves Effects on Earnings

The earnings of immigrants who live in an ethnic/linguistic enclave may differ from the earnings of immigrants who live outside enclave areas. Such an enclaves earnings effect may arise because immigrants are willing to sacrifice some of their earnings to live among others who speak their mother tongue and share their cultural characteristics (ethnic goods). Indeed, for many ethnic goods (e.g., ethnic church, friendship networks, marriage markets, as well as ethnic specific market goods and services), the cost is lower if one lives in a larger ethnic/linguistic enclave (Chiswick and Miller, 1995, 2005b). Thus, only a high wage offer from outside the enclave would induce the immigrant who has a high demand for ethnic goods to live outside the enclave. This gives the appearance of higher nominal wages outside the enclave, although perhaps the same real wages when adjusted for the higher cost of ethnic goods. Other channels of negative influence on earnings of residence in an ethnic enclave include a reduction in job opportunities (Borjas, 2000). Ethnic enclaves can also provide what Borjas (2000) terms a 'warm embrace' that 'gives immigrants information about labour market opportunities, provides many

job contacts, and allows immigrants to escape the discrimination that they may have otherwise encountered in the labour market outside the enclave (Borjas, 2000, p.93). These latter effects would tend to provide immigrants with higher earnings inside the enclave.

Studies of the impact that living in an ethnic enclave may have on immigrants' labor market outcomes have adopted two approaches. Both approaches are based on the expectation that the effect of dominant language proficiency of the individual immigrant on labor market outcomes will vary across birthplace groups that differ in the intensity of the enclave. Under one approach, equations for the determinants of labor market outcomes that include a variable for dominant language proficiency are estimated for various birthplace groups, and the estimated effects of dominant language proficiency are then related to characteristics of the birthplace groups. Under the second approach, an ethnic enclave variable, such as the percentage of the population in the immigrant's area of residence that is from the same country of birth, or that shares the same ethnic or linguistic background, is added to the model of labor market outcomes, and the coefficient on this variable is used in the assessment of the role of ethnic enclaves in the labor market.

Evans' (1987) study for adult male immigrants in Australia in 1981 is an example of the first approach. According to Evans (1987, p. 265) 'The ethnic enclaves hypothesis suggests that the effect of English skills on occupational attainment should be much weaker in groups that have developed ethnic enclaves'. In Australia, immigration from Mediterranean countries has been a consistent feature of post-WWII migration flows, whereas immigration from Eastern European and Third World countries, for the 1981 dataset analysed, was of more recent origin. Consistent with the ethnic enclave hypothesis, English language proficiency had a significantly

larger effect on the occupational status of immigrants from Eastern Europe and Third World countries that it had on the occupational status of immigrants from Mediterranean countries.

Veltman (1983) presented comprehensive analyses of the determinants of occupational status and earnings among adults belonging to a minority language (defined using mother tongue) group in the US, using the 1976 Survey of Income and Education. Separate analyses were undertaken for the Spanish language group and for non-Spanish minority language groups. The language variables included in the estimating equation distinguished four categories: (i) English monolinguals; (ii) workers whose usual language was English and who often speak another language; (iii) minority language speakers who reported that they spoke English either well or very well; and (iv) minority language speakers who reported that they spoke English either not well or not at all. Among the many findings, Veltman reported that among males with a Spanish mother tongue, those whose usual language was Spanish were at an earnings disadvantage, while among those whose usual language was English, the English-Spanish bilinguals were associated with lower earnings. In other words, the labor market 'rewards the most complete type of integration possible, the effective abandonment of the Spanish language as a daily language' (Veltman, 1983, p.241).

In discussing the findings, Veltman (1983, pp.390-391) argues 'As long as Spanish speaking Americans continue to speak Spanish as their usual language, they are relatively well insulated from direct job competition with members of the White English speaking group...As Anglicisation produces increasingly large numbers of English speaking persons of Hispanic ancestry, members of this group are increasingly drawn into competition for jobs with members of the White English language group. It is at this point that the data suggest that Hispanic

Americans are experiencing particular difficulties'. This interpretation is consistent with Evans' ethnic enclaves hypothesis.

Studies that adopt the second, more direct, approach have used two main methodologies. Some studies have augmented a conventional human capital earnings equation with an ethnic concentration variable. In these studies the unit of observation is the individual. Hence, a model of the following type is estimated.

$$\ln Y_i = \alpha_0 + \alpha_1 X_i + \alpha_2 Lang_i + \alpha_3 Conc_i + \varepsilon_i$$

where the vector *X* contains variables for schooling, labor market experience and other factors typically included in studies of the determinants of earnings, *Lang* is a measure of the immigrant's proficiency in the dominant language, and *Conc* is the ethnic concentration variable. This ethnic concentration variable has been defined with reference to the immigrant's ethnic, birthplace or mother tongue group. For example, Shields and Wheatley Price's (2002) analyses in the UK were based on the percentage of the census ward from the same ethnic group as the respondent. In some analyses the dominant language proficiency and ethnic concentration variables are interacted, to enable the effect of the ethnic concentration measure to differ between immigrants who are, and who are not, proficient in the dominant language.

The second method followed in some studies is based around comparisons of the growth in average earnings of groups of immigrants sharing a small number of characteristics. In these studies the unit of observation is the change in a group average, and the estimating equation is of the form

$$\Delta \ln_{jkl} = \alpha_0 + \alpha_1 \ln Y_{jkl}(t_0) + \alpha_2 X_{jkl} + \alpha_3 I_j + \alpha_4 I_k + \alpha_5 I_l + \alpha_6 Conc_{jk} + \varepsilon_{jkl}$$

where $\Delta \ln_{jkl}$ is the growth rate in the mean earnings of immigrant workers from country *j* who live in region *k* and who arrived in the country in year *l*. The initial earnings term, $\ln Y_{jkl}(t_0)$, is used to control for convergence in earnings: the immigrants with the lowest earnings in the first year after arrival experience the most rapid growth in earnings in the destination country (Duleep and Regets, 1997). The vector *X* comprises the proportions of the immigrants in various age and education categories, while I_j, I_k and I_l are vectors of dummy variables that record the *ceteris paribus* earnings differentials across the birthplace, location and year of arrival groups distinguished in the analysis. The *Conc* variable has a construction analogous to that used in the studies that have the individual as the unit of observation. For example, in Borjas (2000) the *Conc* variable is defined as the proportion of the population of metropolitan area *k* $\left(N_k\right)$ who were born in country *j*, namely N_{jk}/N_k.

Examples of studies adopting the first of these approaches are Shields and Wheatley Price (2002) for the UK, Chiswick and Miller (2002) for the US, and Kanas et al. (2012) for Germany.

Chiswick and Miller (2002) based their analysis on the earnings of adult immigrant men in the 1990 US Census. Their concentration variable was defined with respect to languages other than English spoken at home, and was the fraction of the population in the state that speaks the same non-English language at home as the respondent. In the benchmark model, the coefficient of this linguistic concentration variable was negative (-0.006) and highly significant. The coefficient of -0.006 indicates that a one percentage point increase in the linguistic concentration measure would be associated with slightly more than one-half of a percent decrease in an

immigrant's earnings. When a variable for whether the immigrant was proficient in English was added to the model, the coefficient on the linguistic concentration variable changed from -0.006 to -0.005.[29] Further, when the linguistic concentration variable was interacted with the English proficiency variable, the effect of residence in a linguistic concentration for those who were not proficient in English was a statistically significant -0.002, whereas it was -0.006 for those who were proficient in English. Chiswick and Miller also explored the implications of this interaction term from the perspective of how the impact on earnings of proficiency in English varied across regions according to the concentration of speakers of the immigrant's home language. They report (p.43) 'This is estimated to be close to 19% for an individual who lives in an area where his origin language is not spoken. Where 20% of the population speaks the immigrant's origin language, the return to English-speaking skills would be 11%'. In other words, the economic penalty from not speaking English is smaller among those who live in a linguistic concentration area, which supports Evans' ethnic enclaves hypothesis.

Shields and Wheatley Price (2002) examine the determinants of the mean wage in the immigrant's occupation in the UK. Hence, as noted earlier, they capture occupational wage effects but not any intra-occupational wage effects. Their ethnic concentration measure was incorporated into a model of occupational success in the form of three dummy variables, for living in a census ward with 5-15 percent, 15-33 percent, or greater than 33 percent own ethnic density (the benchmark group was 0-5 percent own ethnic density). It was reported that these ethnic concentration variables were only marginally statistically significant, 'and are indicative

[29] The estimated impact of residence in a linguistic enclave in the analyses based on the American Community Survey (2005-2009) reported in Table 4 is also -0.005.

of a 4.5% occupational success penalty to living in a high ethnic, minority density (15%-33%) census ward' (p. 149).

Kanas et al. (2012) examine the role of ethnic concentration among adult male immigrants in Germany using the German Socio-Economic Panel, 1984-2004. They include in their model a number of social capital variables other than the ethnic concentration. The main social capital variable was the frequency of contacts with friends, relatives and neighbors. The models were corrected for sample selection. The results showed that ethnic concentration was not a significant determinant of earnings. Nor was there any evidence that the role of ethnic concentration varied according to the language skills of the immigrant. The authors argue (p.703) 'It is possible that Germany simply lacks substantial ethnic concentrations, like the Cubans in Miami or the Chinese in San Francisco, that significantly influence immigrants' economic outcomes'. Determining whether there are threshold effects in this regard is an area for future research.

There are two main studies that have used the group-average wage-growth approach: Borjas' (2000) analyses for the US and Warman's (2007) more recent study for Canada. Borjas' (2000) analyses were based on the 1980 and 1990 US censuses. His main ethnic concentration measure was the proportion of the population of metropolitan area *k* who were born in country *j* $\left(N_{jk}/N_{k}\right)$. An alternative, relative clustering measure, was also used, where the ethnic concentration measure was normalized by the fraction of the US population that belongs to the particular country of birth group $\left(N_{j}/N\right)$. Borjas reported that the ethnic concentration variable was a negative and statistically significant determinant of earnings growth between 1980 and 1990, and while the results were weaker when the relative clustering measure was used, they

were still indicative of a negative relationship between residential concentration and earnings growth. Borjas reported that the wage growth for a typical Mexican immigrant who moves from Los Angeles (which had an exposure index of 0.11) to New York (exposure index of 0.001) would increase by four percentage points, an effect that was described as sizeable. It was also shown that the adverse impact of residence in an ethnic enclave was stronger for the least educated, and it was also usually stronger for the groups of most recent arrivals, compared to longer-term settlers. In other words, in this study the most disadvantaged in the labor market incurred further penalties from residing in an ethnic enclave.

Warman (2007) examined the impact of residence in an ethnic enclave in the context of earnings growth regressions based on data from the Canadian Census of 1981, 1986, 1991, 1996 and 2001. The dependent variable in the estimating equation was the earnings growth over either 5, 10, 15 or 20 year periods, for the average foreign-born worker from a particular country who resided in a specific Census Metropolitan Area and who belonged to a particular arrival cohort. Along with country of birth, location of residence, arrival cohort and base year fixed effects, the estimating equation also controlled for initial earnings. It was shown that residence in an ethnic enclave has a negative impact on weekly earnings growth, and this negative effect was more important among immigrants who migrated as adults than it was among immigrants who migrated as children, and it was also more important among high-skilled male workers than it was among their low-skilled counterparts. Warman (2007) illustrated the potential impact of the estimates: it was shown that by living in Montreal (low ethnic concentration) rather than in Vancouver (high concentration), an immigrant from Hong Kong would have four percent higher earnings growth over a five-year period. In other words, similar to Borjas' (2000) finding for the

US labor market, the enclave effect is quantitatively important in the Canadian labor market. However, contrary to what Borjas found for the US, Warman's (2007) results indicate that ethnic enclaves afford relative protection to less-skilled workers.[30]

Hence, while the research into the earnings effects of residence in an ethnic enclave is under-developed compared to the other areas of research reviewed in this chapter, the evidence strongly suggests that, even controlling statistically for the respondent's own destination language proficiency, other variables held constant, men who live in an ethnic/linguistic enclave receive, on average, lower earnings than men who live outside of their enclave area.

6. Summary and Conclusions

This chapter reports on the 'economics of language' for immigrants, that is, the influence of language on the choice of destination among international migrants, the determinants of the acquisition among immigrants of destination language proficiency, and the labor market consequences of that proficiency as expressed in their earnings.

There is a tendency among international migrants (and among internal migrants in bilingual countries, such as Canada) to take language issues into account when deciding whether to migrate and the choice of destination. The costs of migration are lower if the migration is to a destination with the same dominant language as that known to the potential migrant. More generally, the smaller the linguistic distance between the dominant languages of the origin and a particular destination, the lower are the adjustment costs in the destination and the higher the

[30] This difference could be due to the different definitions of skilled workers in the studies. Borjas (2000) defines a group as skilled if the mean educational attainment is at least 12 years, whereas Warman (2007) separated the immigrants by highest degree obtained, and defined the skilled group as possessing a university degree.

migration rate.

The determinants of the destination language proficiency among immigrants have been performed primarily for both males and females on census and large surveys for major immigrant receiving countries such as the United States, Canada, Australia, Israel, and Germany. Language proficiency among immigrants is most efficiently modeled according to the three 'Es': Exposure to the destination language in the origin and in the destination, Efficiency in acquiring destination language skills and the Economic incentives for investing in proficiency. Proficiency in the destination language among immigrants increases with the level of their schooling and the duration of their residence in the destination. Proficiency is lower with an older age at migration, if the immigrant was married to the current spouse before migration, and if the migrant lives in a linguistic concentration (enclave) area. Proficiency is also greater the closer are the origin and destination languages (smaller linguistic distance). Among women, but not men, proficiency is lower when there is a larger number of children in the family. There appears to be language learning in the home. The proficiency of a family member is greater if other members of the family are more proficient. In particular, the mother's proficiency is more important than that of the father for the destination language proficiency of their children.

Among immigrants, other variables being the same, earnings are greater for those more proficient in English. The implied payoff to proficiency in terms of labor market earnings for adult males suggests it is a profitable investment. Yet this underestimates the benefits of acquiring proficiency as it does not include the gains from consumption, social and civic activities, and other human capital investments. The computation of benefits also does not take into account the gains from the enhanced English language proficiency of other family members

(language learning in the home) when one family member makes investments in destination language training.

An important implication of this analysis for immigration policy is that immigrants either proficient in the destination language, or with characteristics that enhance proficiency, will be more successful in adjusting to the new labor market. Some countries (such as Australia, Canada the United Kingdom, and New Zealand), but not the United States, have skill-based immigration policies that give significant explicit emphasis to these characteristics, including English language skills (plus French in Canada), educational attainment, occupational skills, and age at migration, when issuing permanent resident visas.

Another important policy implication derives from the high rate of return from investments in language proficiency to the individual and to society. This suggests the encouragement of immigrants to invest in language training, through subsidies, access to training programs and other mechanisms as is done explicitly in some countries, with Israel being a primary example. Encouraging immigrants to become proficient in the destination language does not imply a denigration of their culture or language of origin. It does imply a welcoming of them to the full range of opportunities in the educational, economic, social and civic (political) life of their new home.

The research reviewed in this chapter that has yielded a rich array of findings requires access to quality data. The relative abundance of studies of the determinants of dominant language proficiency, and of the effects of dominant language proficiency on labor market outcomes in Australia, Canada and the US, is a clear reflection of the relative availability in these countries of quality data sets containing information on relevant variables. Conversely, the

limited research for the UK and Spain, until recent years, reflects the absence of the required data in earlier years. The limited amount of recent research into language issues in Israel is a result of the unavailability of data.

The research has revealed the importance of dominant language proficiency to immigrants' labor market outcomes, the differences in these labor market outcomes across groups, the potential differences in labor market outcomes depending on the context in which the destination and origin languages are used, and the changes over time in the economic returns to language usage. More research is needed, however, on the effects of language usage on aspects of the consumer, social, family structure and civic life of immigrants. The inclusion of questions on dominant language proficiency, along with questions on immigrant status and labor market outcomes, demographic and social characteristics, consumption behavior and civic activities in censuses and other large scale data collections, should be viewed as a priority.

References

Adsera, A. and Pytlikova, M., (2012). 'The Role of Language in Shaping International Migration', Institute for the Study of Labor Discussion Paper Number 6333, Bonn, Germany.

Akresh, I.R., (2007). 'Context of English Language Use Among Immigrants to the United States', *International Migration Review*, 41(4), pp.930-955.

Bartel, A.P., (1989). 'Where do the New U.S. Immigrants Live?', *Journal of Labor Economics*, 7(4), pp.371-391.

Bauer, T., Epstein, G.S. and Gang, I.N., (2005). 'Enclaves, Language, and the Location Choice of Migrants', *Journal of Population Economics*, 18(4), pp.649-662.

Becker, G.S., (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education* (1st ed.), New York: National Bureau of Economic Research.

Beine, M., Docquier, F. and Özden, Ç., (2011). 'Diasporas', *Journal of Development Economics*, 95(1), pp.30-41.

Belot, M. and Ederveen, S., (2012). 'Cultural Barriers in Migration Between OECD Countries', *Journal of Population Economics*, 25(3), pp.1077-1105.

Belot, M.V.K. and Hatton, T.J., (2012). 'Immigrant Selection in the OECD', *Scandinavian Journal of Economics*, doi: 10.1111/j.1467-9442.2012.01721.x

Berman, E., Lang, K. and Siniver, E., (2003). 'Language-Skill Complementarity: Returns to Immigrant Language Acquisition', *Labour Economics*, 10(3), pp.265-290.

Bleakley, H. and Chin, A., (2004). 'Language Skills and Earnings: Evidence from Childhood Immigrants', *Review of Economics and Statistics*, 86(2), pp.481-496.

Borjas, G.J., (2000). 'Ethnic Enclaves and Assimilation', Swedish Economic Policy Review, 7, pp.89-122.

Budría, S. and Swedberg, P., (2012). 'The Impact of Language Proficiency on Immigrants' Earnings in Spain', Institute for the Study of Labor Discussion Paper No. 6957, Bonn, Germany.

Carliner, G., (1981). 'Wage Differences by Language Group and the Market for Language Skills in Canada', *Journal of Human Resources*, 16(3), pp.384-399.

Charette, M. and Meng, R., (1994). 'Explaining Language Proficiency: Objective versus Self-

Assessed Measures of Literacy', *Economics Letters*, 44, pp.313-321.

Chiswick, B.R., (1991). 'Speaking, Reading and Earnings among Low-Skilled Immigrants', *Journal of Labor Economics*, 9(2), pp.149-170.

Chiswick, B.R., (1998). 'Hebrew Language Usage: Determinants and Effects on Earnings in Israel', *Journal of Population Economics*, 11(2), pp.253-271.

Chiswick, B.R. (2009). "The Economics of Language Learning for Immigrants: An Introduction and Overview" in T.G. Wiley, J.S. Lee, and R.W. Rumberger (eds.) *The Education of Language Minority Immigrants in the United States*, Bristol, UK: Multilingual Matters, pp. 72-91

Chiswick, B.R., Lee, Y.L. and Miller, P.W., (2003). 'Schooling, Literacy, Numeracy and Labour Market Success', *The Economic Record*, 79(245), pp.165-181.

Chiswick, B.R., Lee, Y.L. and Miller, P.W., (2005a). 'Family Matters: The Role of the Family in Immigrant's Destination Language Skills', *Journal of Population Economics*, 18(4), pp.631-647.

Chiswick, B.R., Lee, Y.L. and Miller, P.W., (2005b). 'Parents and Children Talk: English Language Proficiency within Immigrant Families', *Review of Economics of the Household*, 3(3), pp.243-268.

Chiswick, B.R., Lee, Y.L. and Miller, P.W., (2005c). 'Immigrant Earnings: A Longitudinal Analysis', *Review of Income and Wealth*, 51(4), pp.485-503.

Chiswick, B.R. and Miller, P.W., (1985). 'Immigrant Generation and Income in Australia', *Economic Record*, 61(173), pp.540-553.

Chiswick, B.R. and Miller, P.W., (1988). 'Earnings in Canada: The Roles of Immigrant Generation, French Ethnicity and Language', pp.183-228 in T. Paul Schultz (ed.) *Research in Population Economics*, 6, JAI Press.

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

Chiswick, B.R. and Miller, P.W., (2001). 'A Model of Destination Language Acquisition: Application to Male Immigrants in Canada', *Demography*, 38(3), pp.391-409.

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

Chiswick, B.R. and Miller, P.W., (2003). 'The Complementarity of Language and Other Human Capital: Immigrant Earnings in Canada', *Economics of Education Review*, 22(5), pp.469-480.

Chiswick, B.R. and Miller, P.W., (2005a). 'Linguistic Distance: A Quantitative Measure of the Distance between English and Other Languages', *Journal of Multilingual and Multicultural Development*, 26(1), pp.1-16.

Chiswick, B.R. and Miller, P.W., (2005b). 'Do Enclaves Matter in Immigrant Adjustment?', *City and Community*, 4(1), pp.5-35.

Chiswick, B.R. and Miller, P.W., (2006). 'Language Skills and Immigrant Adjustment: The Role of Immigration Policy', in D.A. Cobb-Clark and S. Khoo (eds). *Public Policy and Immigrant Settlement*, Cheltenham, UK: Edward Elgar, pp.121-148.

Chiswick, B.R. and Miller, P.W., (2007a). *The Economics of Language: International Analyses*, London: Routledge.

Chiswick, B.R. and Miller, P.W., (2007b). 'Computer Usage, Destination Language Proficiency and the Earnings of Natives and Immigrants', *Review of Economics of the Household*, 5(2), pp.129-157.

Chiswick, B.R. and Miller, P.W., (2008a). 'A Test of the Critical Period Hypothesis for Language Learning', *Journal of Multilingual and Multicultural Development*, 29(1), pp.16-29.

Chiswick, B.R. and Miller, P.W., (2008b). 'Modeling Immigrants' Language Proficiency', in B.R. Chiswick (ed.) *Immigration: Trends, Consequences, and Prospects for the United* States, Amsterdam: Elsevier, pp.75-128.

Chiswick, B.R. and Miller, P.W. (2012). 'The Impact of Surplus Skills on Earnings: Extending the Over-Education Model to Language Proficiency', Xerox.

Chiswick, B.R., Patrinos, H.A. and Hurst, M.E., (2000). 'Indigenous Language Skills and the Labor Market in a Developing Economy', *Economic Development and Cultural Change*, 48(2), pp.349-367.

Chiswick, B.R. and Repetto, G., (2001). 'Immigrant Adjustment in Israel: The Determinants of Literacy and Fluency in Hebrew and their Effects on Earnings', in S. Djajic (ed.) *International Migration: Trends, Policies, and Economic Impact*, London: Routledge. pp.204-228.

Christofides, L. and Swidinsky, R., (2010). 'The Economic Returns to the Knowledge and Use of a Second Official Language: English in Quebec and French in the Rest-Of-Canada',

Canadian Public Policy, 36(2), pp.137-158.

Clark, X., Hatton, T.J. and Williamson, J.G., (2007). 'Explaining U.S. Immigration, 1971-1998', *Review of Economics and Statistics*, 89(2), pp.359-373.

Cobb-Clark, D.A. and Khoo, S., (2006). *Public Policy and Immigrant Settlement*, edited volume, Cheltenham, UK: Edward Elgar.

Crystal, D., (1987). *The Cambridge Encyclopedia of Language*, Cambridge: Cambridge University Press.

Duleep, H.O. and Regets, M.C., (1997). 'Immigrant Entry Earnings and Human Capital Growth: Evidence from the 1960-1980 Censuses', *Research in Labor Economics*, 16, pp.297-317.

Dustmann, C., (1994). 'Speaking Fluency, Writing Fluency and Earnings of Migrants', *Journal of Population Economics*, 7(2), pp.133-156.

Dustmann, C., (1999). 'Temporary Migration, Human Capital, and Language Fluency of Migrants', *Scandinavian Journal of Economics*, 101(2), pp.297-314.

Dustmann, C. and Fabbri, F., (2003). 'Language Proficiency and Labour Market Performance of Immigrants in the UK', *Economic Journal*, 113(489), pp.695-717.

Dustmann, C. and van Soest, A.V., (2001). 'Language Fluency and Earnings Estimation with Misclassified Language Indicators', *Review of Economics and Statistics*, 83(4), pp.663-674.

Dustmann, C. and van Soest, A.V., (2002). 'Language and the Earnings of Immigrants', *Industrial and Labor Relations Review*, 55(3): 473-492.

Dyen, I., Kruskal J.B. and Black, P., (1992). An IndoEuropean Classification: A Lexicostatistical Experiment, *Transactions of the American Philosophical Society*, New Series 82(5).

Espenshade, T. and Fu, H., (1997). 'An Analysis of English Language Proficiency Among U.S. Immigrants', *American Sociological Review*, 62(2), pp.288-305.

Espinosa, K.E. and Massey, D.S., (1997). 'Determinants of English Proficiency among Mexican Migrants to the United States', *International Migration Review*, 31(1), pp.28-50.

Evans, M.D.R., (1986). 'Sources of Immigrants' Language Proficiency: Australian Results with Comparisons to the Federal Republic of Germany and the United States of America', *European Sociological Review*, 2(3), pp.1-10.

Evans, M.D.R., (1987). 'Language Skill, Language Usage and Opportunity: Immigrants in the

Australian Labour Market', *Sociology*, 21(2), pp.253-274.

Ferrer, A., Green, D.A. and Riddell, W.C., (2006). 'The Effect of Literacy on Immigrant Earnings', *Journal of Human Resources*, 41(2), pp.380-410.

Ginsburgh, V., Ortuño-Ortín, I. and Weber, S., (2005). 'Disenfranchisement in Linguistically Diverse Societies. The Case of the European Union', mimeo, Department of Economics, University of Alicante, Spain.

Grenier, G., (1984). 'Shifts to English as Usual Language by Americans of Spanish Mother Tongue', *Social Science Quarterly*, 65(2), pp.537-550.

Grenier, G., (1987). 'Earnings by Language Group in Quebec in 1980 and Emigration from Quebec between 1976 and 1981', *Canadian Journal of Economics*, 20(4), pp.774-791.

Grogger, J. and Hanson, G.H., (2011). 'Income Maximization and the Selection and Sorting of International Immigrants', *Journal of Development Economics*, 95(1), pp. 42-57.

Hart-Gonzalez, L. and Lindermann, S., (1993). 'Expected Achievement in Speaking Proficiency, 1993', School of Language Studies, Foreign Service Institute, U.S. Department of State, Mimeographed.

Hou, F., (2005). 'The Initial Destinations and Redistribution of Canada's Major Immigrant Groups: Changes over the Past Two Decades', Statistics Canada Research, Catalogue No. 11F0019MIE—No. 254.

Hutchinson, W.K., (2005). '"Linguistic Distance"' as a Determinant of Bilateral Trade', *Southern Economic Journal*, 72(1), pp.1-15.

Isphording, I.E. and Otten, S., (2011). 'Linguistic Distance and the Language Fluency of Immigrants', RUHR Economic Papers #274, Bochum, Germany.

Isphording, I.E. and Otten, S., (2012). 'The Costs of Babylon—Linguistic Distance in Applied Economics', RUHR Economic Papers #337, Bochum, Germany.

Jaeger, D.A., (2000). 'Local Labor Markets, Admission Categories, and Immigrant Location Choice, unpublished manuscript, Department of Economics, College of William and Mary, USA.

Kan, A., (1991). *The Australian Experience of Skilled Migration: The Employer Nomination Scheme in the Manufacturing and the Finance, Property and Business Services Sectors*, Bureau of Immigration Research, Australian Government Publishing Service, Canberra, Australia.

Kanas, A., Chiswick, B.R., van der Lippe, T. and van Tubergen, F., (2012). 'Social Contacts and the Economic Performance of Immigrants: A Panel Study of Immigrants in Germany', *International Migration Review*, 46(3), pp.680-709.

Karemera, D., Oguledo, V.I. and Davis, B., (2000). 'A Gravity Model Analysis of International Migration to North America', *Applied Economics*, 32(13), pp.1745-1755.

Kominski, R., (1989). 'How Good is "How Well"? An Examination of the Census English-Speaking Ability Question', *American Statistical Association 1989 Proceedings of the Social Statistics Section*, pp.333-338.

Kossoudji, S.A., (1988). 'English Language Ability and the Labor Market Opportunities of Hispanic and East Asian Immigrant Men', *Journal of Labor Economics*, 6(2), pp.205-228.

Lazear, E., (1999). 'Culture and Language', *Journal of Political Economy*, 107(6), pp.595-612.

Lewis, M.P., (2009). *Ethnologue: Languages of the World*, Sixteenth edition. Dallas, Texas: SIL International.

Long, M.H., (1990). 'Maturational Constraints on Language Development', *Studies in Second Language Acquisition*, 12, pp.251-285.

Marschak, J., (1965). 'Economics of Language', *Behavioral Science*, 10, pp.135-40.

Mayda, A.M., (2010). 'International Migration: A Panel Data Analysis of the Determinants of Bilateral Flows', *Journal of Population Economics*, 23(4), pp.1249-1274.

McAllister, I., (1986). 'Speaking the Language: Language Maintenance and English Proficiency among Immigrant Youth in Australia', *Ethnic and Racial Studies*, 9(1), pp.24-42.

McCloskey, D.N., (1998). *The Rhetoric of Economics*, 2nd edition, The University of Wisconsin Press, Madison, US.

McDonald, J.T., (2003). 'Location Choice of New Immigrants to Canada: The Role of Ethnic Networks', in C.M. Beach, A.G. Greene and J.G. Reitz (eds.) *Canadian Immigration Policy for the 21st Century*, John Deutsch Institute for the Study of Economic Policy, Queen's University, Canada.

McDonald, J.T., (2004). 'Toronto and Vancouver Bound: The Location Choice of New Canadian Immigrants', *Canadian Journal of Urban Research*, 13(1), pp.85-101.

McManus, W., Gould, W. and Welch, F., (1983). 'Earnings of Hispanic Men: The Role of English Language Proficiency', *Journal of Labor Economics*, 1(2), pp.101-130.

Miranda, A. and Zhu, Y., (2013). 'English Deficiency and the Native-Immigrant Wage Gap', *Economics Letters*, 118(1), pp.38-41.

Nadeau, S. and Seckin, A., (2010). 'The Immigrant Wage Gap in Canada: Quebec and the Rest of Canada', *Canadian Public Policy*, 36(3), pp.265-285.

Ortega, F. and Giovanni, P., (2009). 'The Causes and Effects of International Migrations: Evidence from OECD Countries 1980-2005', Working Paper No. 09.6, Department of Economics, University of California, USA.

Pedersen, P.J., Pytlikova, M. and Smith, N., (2008). 'Selection and Network Effects-Migration Flows into OECD Countries 1990-2000', *European Economic Review*, 52(7), pp.1160-1186.

Raijman, R., (2013). 'Linguistic Assimilation of First-Generation Jewish South African Immigrants in Israel', *Journal of International Migration and Integration*, forthcoming.

Schultz, T.W., (1962). 'Investment in Human Beings', Special Supplement to *Journal of Political Economy*, 70(5), pp.1-157.

Shields, M.A. and Wheatley Price, S., (2002). 'The English Language Fluency and Occupational Success of Ethnic Minority Immigrant Men Living in English Metropolitan Areas', *Journal of Population Economics*, 15(1), pp.137-160.

Stevens, G., (1986). 'Sex Differences in Language Shift in the United States', *Social Science Research*, 71(1), pp.31-36.

Tainer, E., (1988). 'English Language Proficiency and Earnings among Foreign-Born Men', *Journal of Human Resources*, 23(1), pp.108-122.

Tanakh: The Holy Scriptures (1985). Philadelphia: Jewish Publication Society.

Veltman, C.J., (1983). *Language Shift in the United States*, New York: Mouton.

Warman, C., (2007). 'Ethnic Enclaves and Immigrant Earnings Growth', *Canadian Journal of Economics*, 40(3), pp.401-422.

Zavodny, M., (1999). 'Determinants of Recent Immigrants' Location Choices', *International Migration Review*, 33(4), pp.1014-1030.