The effect of linguistic proximity on the labour market performance of immigrant men in Canada

Alícia Adserà
Woodrow Wilson School
Princeton University
347 Wallace Hall,
Princeton NJ 08544
adsera@princeton.edu

Ana Ferrer
Department of Economics
University of Waterloo
200 University Avenue
Waterloo, ON, Canada N2L 3G1
aferrer@uwaterloo.ca

Abstract

We argue that the ability to speak the language of the destination country plays a key role in the labour market performance of immigrant men, particularly the university educated. To that end, we combine large samples of the restricted version of the Canadian Census (1991-2006) with a new measure of linguistic proximity of the *most-used* language in the immigrant's country of origin to that of the destination country, and with information about the occupational skills required for the jobs the immigrant holds. Immigrants with more distant languages earn lower wages and work in jobs requiring more physical strength and fewer social and analytical skills than the jobs of similar native-born workers. More importantly, linguistic distance imposes a relatively large wage penalty to the university-educated upon entry to the country. However, both the wage and the analytical skill requirements of the jobs held by university-educated immigrant men increase, while the strength skill requirements decline, with time in the country. Our analysis suggests that facilitating language acquisition may speed up the labour market assimilation of immigrant men in Canada.

Key Words: skill and wage assimilation; labour market outcomes of immigrant, wage gaps, Canadian migration

JEL: J15, J24, J31, J61, F22

The analysis was conducted at the South Western Ontario RDC which is part of the Canadian Research Data Centre Network (CRDCN). The services and activities provided by the CRDCN are made possible by the financial or in-kind support of the SSHRC, the CIHR, the CFI, Statistics Canada and participating universities whose support is gratefully acknowledged. The views expressed in this paper do not necessarily represent the CRDCN's or those of its partners'.

1. Introduction

Over the past two decades, about 60 percent of all immigrants to Canada entered the country as economic immigrants and were assessed through a points system that rewards general human capital, most notably education.[1] However, despite the explicit emphasis of Canadian immigration policy on education, both the earnings and the employment outcomes of recent immigrant cohorts arriving to Canada have experienced a significant deterioration.[2] In this paper, we explore the role of language in the assimilation of immigrant men to Canada, particularly the university-educated.

Although linguistic proficiency has always been embedded in the Canadian point system, poor linguistic ability has often been invoked as one of the reasons behind observed differences in outcomes between immigrant and native-born workers.[3] However, the effect of linguistic proficiency on labour market performance of Canadian immigrants is not well understood, because the Canadian census lacks a proper measure of linguistic ability. The few studies that have tried to address this issue typically rely on self-reported measures, such as those provided in the Census or in the Longitudinal Survey of Immigrants to Canada (LSIC). However, self-reported measures can be problematic because they can be affected by shifts in optimism, when migrants assess their own linguistic ability as they spend more time in the host country.[4] Further, the survey questions may be too general to produce meaningful variance in linguistic ability. The Canadian Census linguistic proficiency question, which asks respondents whether they speak either English and/or French well enough to conduct a conversation, receives over 90% positive responses.

---

[1] This system was introduced in the mid-1960s to respond to short-term labour market needs, using points to prioritize specific occupations. For a thorough description of the immigration system, see Ferrer et al. (2014) and references therein.

[2] See, e.g., Baker and Benjamin (1994) and Aydemir and Skuterud (2005). This has largely been attributed to changes in immigrants' background or to economic conditions at the time of entry (Hou and Picot, 2009). The role of differences in educational quality and the low return to foreign human capital have also been explicitly considered in Li and Sweetman (2014), Ferrer and Riddell (2008), Ferrer et al (2006) and Green and Worswick (2012). More recently, Warman and Worswick (2015) emphasize the role of skill-biased technical change in the decline of returns to immigrants. A similar trend among immigrants in the U.S. has been documented (e.g., Borjas 1995, Friedberg, 2000; Bratsberg and Raaum, 2004, Borjas and Friedberg 2009).

[3] Pre-IRPA, proof of language ability was subject to interviewer's approval. Only after the introduction of IRPA in 2002, applicants must submit either written evidence demonstrating their language ability in Canada's official languages or provide language test results from an approved organization or institution, although the test *per se* is not mandatory (Begin et al., 2010).

[4] This issue has been reported to be the case in the LSIC and the German GSOEP

In this paper we use a new measure of linguistic proximity to proxy the influence of language fluency on the labour market assimilation of immigrant men in Canada. Moreover, we focus on the evolution of occupational skills in the jobs held by immigrant men as well as their wages with time in the country. We find that immigrant men who speak more distant languages earn lower wages and work in jobs requiring relatively more physical strength and fewer social and analytical skills upon arrival. This is particularly the case for university-educated immigrants. Interestingly, however, we also find that the speed of assimilation of university-educated immigrant men in Canada is positively related to language distance to English. In particular, we document a significant decrease in the strength skill content, and an increase in the analytical skill content of the jobs held by university-educated immigrant men with time in the country. Our findings suggest that language training programs for recent immigrants may facilitate the labour market assimilation of immigrant men in Canada.

Recent studies (Chiswick and Miller, 2005; Isphording and Otten, 2011) have validated the use of linguistic proximity as an adequate proxy for linguistic ability, showing that it is easier for a foreigner to acquire a language if his native language is linguistically close to the language to be learned. Our measure of linguistic proximity, which was developed in Adserà and Pytlikova (2015), measures the distance in the linguistic tree between origin and destination languages, to approximate linguistic ability. With our new measure, we are able to uncover much richer patterns in the role of language on the assimilation of Canadian immigrants than can be derived from a single dummy indicating common language. Further, compared to self-reported measures, it provides an objective measure of linguistic distance removed from individual. We link this measure to men's information from the confidential Canadian Micro Census files from 1991-2006 to track the labour market performance of different arrival cohorts. Therefore, our study is able to assess the effect of language on the *changes* in labour market outcomes over the medium to long term for different levels of linguistic proximity.

Moreover, our paper contributes to the literature on occupational skills by examining the effect that linguistic proximity might have on the career progression of immigrants, measured by changes in the level of skills required in the jobs they hold. The nature of the work performed in a particular job is a dimension along which the labour force outcomes of immigrant and native-born workers are likely to differ. If immigrants have relatively more difficulty in finding jobs that match their skills, due to poor language proficiency, they may end up in more unskilled jobs than similar native-born. The extent to which this under-placement becomes permanent rather than temporary is a matter of great policy concern.

The study of occupational skills, pioneered by Autor (2013), is particularly relevant to understand the dynamics of immigrant assimilation. Considering broad measures of occupational status – such as blue/white collar or managerial/non-managerial jobs – to analyze these dynamics may miss a substantial part of the heterogeneity within them, but including too finely detailed 4-digit level SOC occupations in the econometric specification is impractical. Imai et al. (2014) follow Autor et al (2003) in using the detailed information contained in occupation databases, such as the Dictionary of Occupational Titles (DOT) or the Occupational Information Network (O*NET), to derive a small set of fundamental skill requirements (physical strength, analytical or social) for each job. These measures have the advantage of being limited in range, but account for variation across very detailed occupational classifications. We borrow these measures to perform our analysis on the occupational assimilation of immigrant men.

To the best of our knowledge, ours is the first paper to specifically examine the effect of language proficiency on changes in occupational skills over time in Canada. Berman et al. (2003) show the existence of complementarities between language proficiency and occupations in a limited set of occupations. Imai et al. (2014) explore the role of language proficiency in the transfer of occupational skills from the home to the host country for a sample of immigrants arriving to Canada in 2001. By contrast, the use of four Census years allows us to study all recent cohorts since the late 1980s (rather than a single cohort) and track them across time. Moreover, Census data allows a policy-relevant comparison between immigrants' labor market performance and that of the native-born. Finally, our study uses an objective measure of linguistic proximity that does not rely on self-report and provides a wide gradient of proficiency to uncover a richer picture of the effect of language on labour market outcomes.

We find that, upon arrival, weekly wages are lower for men from more linguistically distant countries and that their jobs involve relatively more physical strength and less social and analytical skills than those of similar natives. This effect is particularly strong for university-educated immigrants. More interestingly, when we examine their assimilation patterns, we find that, with time in the country, wages increases substantially for university-educated immigrant men, while analytical skills requirements of their jobs increase and the strength requirements decline.

To the extent that immigrant men are positively selected along their unobserved linguistic fluency, our results may under-estimate the true effect of fluency. Further, since linguistic distance may be correlated with other confounders at the country level (such as ethnicity) we conduct a series of robustness exercises by including controls for region of birth, genetic distance between origin and destination, and whether English is the mother tongue of the immigrant. In addition, we show that the linguistic proximity of the father's country of birth does not explain the labour market performance of their second generation children, as expected if our measure is a good proxy of fluency.

The rest of the paper is organized as follows. Section 2 describes the data and the linguistic proximity measure used in the analysis in detail. Sections 3 and 4 present results for weekly wages and skill requirements, first for all men and then separately by educational attainment. Section 5 discusses the robustness of findings to the use of a more generous measure of linguistic proximity. Section 6 concludes.

2. Data and Methodology

2.1. Sample description

Data comes from the restricted files of the Canadian Census (1991, 1996, 2001 and 2006). The confidential files offer detailed information on individual fertility, occupation, country of birth, year of arrival, and mother tongue for 20% of the Canadian population.

The sample includes men aged 18 to 60.[5] In order to reduce computing time to a reasonable length, we select all immigrants (who arrived at age 18 or older) plus a 25 percent random sample of Canadian-born individuals from each census, which provides a sufficiently large sample of the reference group. Observations are weighted accordingly. The sample only includes immigrants who arrived as adults because their behaviour and language ability are quite different from those of immigrants arriving as children or teens (Shaafsma and Sweetman, 2001). In addition, each immigrant record is linked to information about his country of birth at the time of migration including GDP per capita, fraction of individuals with higher education and female labour force participation obtained from Frank and Hou (2013). Summary statistics in Table 1 show that

[5] Adsera and Ferrer (2014, 2016) analyze the occupational mobility of immigrant women and their wages.

immigrants are generally older than the native born, and more likely to have university education and be married than natives.

2.2. Linguistic Proximity

Previous research in the US, Europe and Australia suggests that both proficiency in the language of the destination country and/or the ability to learn it quickly influence an immigrant's success in the destination labour market.[6] The Canadian Census includes a self-reported variable of language ability that does not offer much variation since 91.1% of our immigrant sample reports being able to conduct a conversation in English. In absence of direct measures of linguistic ability, we use indicators developed by Adserà and Pytliková (2015) that vary according to how many levels of the linguistic family tree the languages of both the destination and the source country share. The metric is based on information from the encyclopaedia of languages Ethnologue (Lewis, 2009) a comprehensive catalogue of living languages that situates each language within its family in the linguistic tree. A significant advantage of this metric is that it avoids the endogeneity and measurement error problems associated with self-reported measures of linguistic fluency.[7]

We match the linguistic proximity (LP) indicators to individuals through the variable "place of birth" in the Canadian Census that specifies the person's country of birth. Our metric considers the distance between the *most used* language at origin and the most used language at destination (English). A first indicator of linguistic proximity equals one when two languages do not share any level of the linguistic tree (Chinese and English) and zero otherwise. The second one denotes that two languages are only related at the first most aggregated level of the linguistic tree. For example, the indicator equals one for all pairs of Indo-European languages, such as Hindi and English, which do not share any other branch in the tree. The third and fourth linguistic proximity dummies indicate that two languages only belong to the same first and second-level of the linguistic tree (English and Danish) or up to the third level (English and German). One additional

[6] Most of the US studies use self-reported measures of fluency (Kossoudji 1988; Chiswick and Miller 2002, 2010) and find a negative association between lack of fluency and wages or occupational status. Similar work has also been done in Europe (Dustmann, 1994; Dustmann and van Soest, 2002; and Dustmann and Fabbri, 2003) and Australia (Chiswick and Miller, 2007) looking both at earnings and employment.

[7] Linguistic proximity is also important for the selection of destination since migrants themselves recognize its importance for their subsequent adaptation in the labour market and society as a whole (Adserà and Pytlikova 2015)

indicator equals one if both languages are the same (or share at least the fifth level in the tree which makes them exceptionally close).[8]

Table 1 shows the fraction of immigrants that fall into each LP category. Around 48% of immigrant men come from a country where the most widely-spoken language has no common branch with English. Table 2 reports the main source countries in each category of LP according to the *most-used* metric. Percentages in the table denote the share of immigrants coming from a particular country among all the countries that share the same distance to English. In the first row for example, Chinese migrants constitute 26% of all migrants whose most used language does not share any branch of the linguistic tree with English. The other main source countries in this category are Philippines, Vietnam, Pakistan, Jamaica and Korea, those where the first official language is Arab (Egypt, United Arab Emirates) and most African countries. Countries sharing only one branch in common (Indo-European languages) comprise 36% of immigrants. Immigrants from Northern European countries (Denmark, Iceland, Norway and Sweden), who represent just 0.004% of immigrants to Canada, share two levels of the linguistic tree with English. Countries of Germanic origin (Switzerland, Germany, Belgium, The Netherlands and their colonies) have three linguistic branches in common with English. This is also a relatively small group, accounting for 3% of immigrants. Finally, the main countries whose most used language is English are US, UK, Ireland, as well as some former colonies such as Trinidad. This group comprises 13% of the immigrants in the sample.

The set of LP variables provides a far more detailed metric of proximity between any pair of languages than the standard indicator for common language used in most of the literature. Further, while still being closely correlated to the self-reported measure in the Census, they capture better the heterogeneity of origins to proxy language ability. The percentages of immigrants who share none or only one branch in common with English and report not being able to hold a conversation in English are 41% and 56% respectively. The same percentage stands at 2% and 0% for those who share three branches or have English as the major spoken language in their countries. Further, self-reported assessments suffer from endogeneity problems and measurement error (Dustmann and van Soest (2001). In contrast the linguistic proximity metric is time invariant, and allows the

[8] Under this metric there is no indicator for the 4th-level linguistic proximity because no migrant origin language shares only four branches with English, as happens among Romance languages such as French, Italian and Spanish.

effect of language-acquisition in labour performance to be picked up through measures of time spent in Canada.

The "*most-used*" index likely indicates quite precisely (at least statistically) the linguistic ability of the average citizen in the country, as it picks up proximity between the language most commonly used at origin and English. However, immigrants are a selected group of the population who can arrive either as refugees escaping economic or political hardship, as economic immigrants seeking better opportunities, or as relatives of settled immigrants. Therefore, to the extent that selection is moderate, the proposed measure might be appropriate. In order to consider alternative possibilities in terms of immigrant selection, section 5 employs instead the minimum distance between "*any official*" or "*most-used*" languages at origin and destination.[9]

We also employ an alternative measure of linguistic distance, the Levenshtein index, which relies on phonetic dissimilarity of the 40 most common words between each pair of all world languages (see Bakker et al. 2009). It has been previously used by Isphording and Otten (2011) as a good predictor of language proficiency and by Adserà and Pytlikova (2015) to explain the choice of migrant destination. In Table 1 the index (that ranges from 0 for equal languages to around 106) has an average of 82.6 among migrants in the sample.

A natural concern when using proxy variables for linguistic proficiency is that they may be picking up to some extent other country of origin confounders that can affect immigrant performance. In particular, either the quality or just misconceptions about the quality of education in the country of origin can result in lower returns to foreign education (Ferrer and Riddell 2008, Li and Sweetman, 2014).[10] Unfortunately, given the limited variation in the measure of self-reported language ability in the Census, we cannot use our linguistic proximity index as an instrument for that measure as others in the literature have done. To circumvent this limitation and minimize confounding effects, our analysis controls for a variety of country of origin

---

[9] We also consider a third metric that measures the distance between the first official language at origin and destination, denoted *"first official*". Compared to the *most used* language, this metric involves a major change for immigrants from South Africa, Jamaica and Guyana, who do not share any branch with English under the *most used* criteria, but share 3 branches (South Africa) or use the same language (Jamaica and Guyana) under the *first official* language metric. However, since results are essentially the same with both metrics, we only report those using the *most used* language to keep the paper at a reasonable length.

[10] Work on the effect of credentials on immigrant labour market outcomes suggests that lack of credentials imposes a penalty on immigrant wages, although not to the same extent that it does for the native-born (Ferrer and Riddell (2008).

characteristics that proxy for level of development and institutional quality as well as broad region of birth indicators. More importantly, a worker's ethnicity, that has been shown to impose barriers to employment and to hinder career progress in Canada, could be conceivably confounded with limited language proficiency (Oreopolous, 2011). To rule out this possibility we devise a robustness exercise where we look at the effect that the linguistic proximity of immigrants' fathers has on the wages of their Canadian born children. If our measures are good proxies of linguistic proximity, they should not be related to the performance of the second generation.

2.3. Labour Outcomes: Wages and Skill-requirements

The Census collects information on the wages and weeks worked the previous year; however, information on hours of work is collected for the census reference week. For this reason Canadian studies using Census data use weekly wages – rather than hourly wages – as a measure of labour market performance.[11] We follow this strategy as well, which renders our results comparable to previous Canadian estimates on the performance of immigrants.

As a second measure of performance, we employ a set of indexes developed by Imai et al (2014) to measure the skill requirement of jobs in the Canadian market. The rationale behind this exercise is that many job changes that involve significant adjustments in the skills employed in the workplace may not result in shifts in broad occupational categories. In order to capture variation across narrowly defined occupations without overwhelming estimation procedures, Imai et al (2014) follow Autor et al. (2003) and Autor and Handel's (2013) methodology and use the detailed information in the Occupational Information Network (O*NET) to construct a low-dimensional vector of skill requirements that account for variation in very detailed occupational classifications at the 4-digit occupational category using Principal Component Analysis (PCA). In particular, the vector includes two indexes for cognitive skills (social and analytical) and three indexes for manual skills (fine motor skills, physical strength, and visual skills). This classification differs from that employed by those who study the routine/non-routine nature of the jobs (Autor and Dorn, 2013; Warman and Worswick, 2015) to understand the effect of skill-biased technological change on wage inequality. We abstract from this distinction to focus on a more general notion of manual and non-manual skills.

---

[11] See Warman and Worswick (2015), Ferrer and Riddell (2008), and Aydemir and Skuterud (2005), among others.

We link these indexes to each individual's employment record through 4-digit occupational codes available in the Census. Note that these indexes measure skills from the demand side of the market, i.e. the skills involved in performing each job, not the actual skills of the worker. The indexes are easily interpretable: a positive value for the analytical index indicates that a job demands more analytical skills than the job held by the average worker. Further, one unit of the skill score (with mean zero) is equivalent to one standard deviation in the skill distribution of the Canadian population.[12]

The correlation between skills is substantial. Strength skills are positively and significantly correlated with motor and visual skills, and negatively correlated with analytical and social skills (see Table 1A in the Appendix). The latter two are positively and strongly correlated between themselves. Although we uncover significant differences across all skills in the analyses, we limit the discussion in the paper to the three that appear more relevant. Table 1 shows that immigrant men hold jobs that require, on average, similar interpersonal skills, higher analytical skills and lower strength skills than the average Canadian-born male. A look at the full distribution of skills (Figure 1) unveils that immigrant men work in jobs that are more heavily concentrated in the lower tail of the strength distribution and that the distribution of their required analytical and social skills is flatter than that of the native born, with lower fractions in the middle.

3. Labour market outcomes and Linguistic Proximity

3.1. The effect of linguistic proximity on wage and skill differentials

Our basic specification is a common Mincer earnings equation augmented to consider the role of linguistic proximity between origin and destination languages (a proxy of language proficiency).

$$\log(w \quad) = X\beta_1 + +\beta_2 \sum_l \; L_l \quad + \beta_3 \sum_t \; year_t + \epsilon \qquad (1)$$

The dependent variable (*log w*) is a measure of the log weekly wages of individual *i*. We omit this sub-index to avoid cluttering. The vector X contains standard individual demographic characteristics such as marital status of residence, and Census Metropolitan Area), education, experience and experience squared, as well as characteristics of the country of birth at the time of

[12] The principal component analysis in Imai et al. (2014) calculates factor loadings so that they maximize the variation of the data explained by index. The procedure uses as weights the distribution of skills of the Canadian working population. A detailed description of the procedure can be found in their paper.

immigration such as GDP per capita. We measure the difference in earnings of migrants with respect to natives (the reference group) with five indicators of immigrant linguistic proximity L_l. These range from "no linguistic proximity" (for origin countries whose major language does not share any branch with English) to an indicator for "same language" (for origin countries where English is the major language). A set of year dummies, *t = 1995, 2000 and 2005*, controls for the year when weekly wages of the individual were measured (1990 is the omitted category).

Table 3 reports a set of estimates from equation (1) that show the relation between the linguistic proximity of the major language in an immigrant's country of origin to English and his log weekly wages relative to a similar native-born individual, after controlling for basic demographic variables. These models do not include any indicator of years since migration because, rather than focusing on the differences of wages at *entry* by linguistic proximity, this first specification aims to capture average differences across groups.[13] We explore the variation with time since migration to Canada in a fully flexible specification below.

As expected, immigrants with closer linguistic proximity to English have higher weekly wages than those with lower proximity. Column 1 shows that immigrants whose languages share no branch in common with English receive weekly wages around 38% lower than otherwise similar native-born workers. The wage penalty gradually diminishes as languages become closer to English, with the exception of immigrants from countries with a 2nd level of linguistic proximity (corresponding to North European languages such as Danish or Swedish), who experience lower penalties (around 6%) than those who share three levels (with 22% lower wages than natives). Those immigrants arriving from countries where English is the major language only experience a 5% penalty. These estimates are consistent with those derived from other measures used in the literature (Chiswick, B. and P. Miller (2015). Since all models in Table 3 include GDP per capita in the country of origin at the time of migration, these estimated differentials are net of the level of development in the country of origin, which is likely correlated with the quality of the education received by migrants in their source countries -a potential source of wage differentials

The second column shows estimates from a similar model that employs the Levenshtein index of linguistic distance as an alternative proxy for fluency. The coefficient for the first

[13] When we introduce a control for *years since migration* in specifications similar to those in columns (1) and (3) in Table 3, we obtain estimates for the value of a given level of LP *at entry* that are consistent with those shown in Table 3 – they show similar ranking but smaller coefficients. These estimates are available upon request.

generation immigrant indicator measures the wage gap between migrants arriving from countries with English as major language (and a Levenshtein indicator equal to zero) and the native-born. The estimated negative coefficient for the Levenshtein index entails a larger penalty for those more distant to English. Results in column (2) imply that weekly wages of an English speaker migrant are 4% lower than those of a native born. Earnings of an immigrant with the average linguistic distance in the sample (a Levneshtein score 80) are around 24% lower than those of an immigrant from an English-speaking country. An immigrant with the widest distance to English (an index around 106) earns 36% less than natives. [14] These wage penalties are in line with those in column (1) where the indicators from the Adserà and Pytliková (2015) linguistic proximity index are used instead.

As discussed above, linguistic proximity can be associated with other traits such as ethnicity, religion, the quality of educational institutions in the country of origin, or even the actual level of exposure to the foreign language through mass media. The third column of Table 3 introduces controls for place of birth (POB), employing broad geographic regions, with the purpose to better differentiate linguistic proficiency from other omitted characteristics associated with country of birth. We cannot employ country of birth fixed effects because they would be perfectly correlated with the LP indicators. The estimated wage differentials in column (3) shrink substantially for some linguistic groups compared to those in column (1), suggesting that LP partially picks up the effect of other confounding characteristics associated with country of origin. In particular, the wage differential for immigrants whose language has no connection to English with respect to natives decreases from 38% to 21%. The gap also closes for those with either one or three branches in common with English and it is no longer significant for immigrants with a 2nd level LP (Northern European countries).[15]

To further evaluate to what extent our linguistic proximity variable is picking up language ability, we conduct three additional exercises. First, we introduce a measure of "dominant" genetic distance between the population of the country of origin and Canada, to address the concern that the LP measure is related to immigrant ethnicity or culture. The index, based on the work by

[14] Estimates employing the Dyen index that measures similarity only between Indo-European languages produced similar results and are available upon request.

[15] This result for Nordic languages is driven by a combination of two factors: first, the small number of immigrants in this category, and second the fact that more than 80% of the population in Scandinavian countries speaks fluent English (see also footnote 14).

Cavalli-Sforza, Menozzi, and Piazza (1994), increases with the distance between the largest ethnic groups in the two countries and equals zero if the distributions of alleles in both populations are identical. It has already been used in studies of development, linguistic ability or international migration (Spolaore and Wacziarg 2009, Isphording and Otten (2014), Adserà and Pytlikova 2015). Column 4 shows that the coefficients for the LP indicators remain virtually unchanged when this measure is included jointly with the POB.[16]

In a similar spirit, the model in column (5) includes an indicator that equals one when English is the mother tongue of an immigrant. This offers us another way of measuring whether our LP index accurately measures English proficiency. The coefficient for this indicator is positive and denotes a reduction of 10 percent of the wage difference for those migrants whose mother tongue is English regardless of the LP associated to their country of origin. As expected, this addition significantly changes the coefficient of our indicator for "same language", suggesting that this group may be composed both of migrants whose mother tongue is English and of some individuals who arrived from countries where the most used language is English, but did not learn it as children. Note that this is not conclusive evidence that the latter group of immigrants is not fluent in English, although the larger negative coefficient for the "same language" indicator in column (5) of -0.12 as compared to -0.02 in column (3) indicates a disadvantage for this group. The net wage gap for migrants who spoke English since childhood in those origins stays at 2 per cent as in column (3). All other coefficients remain similar, with only minor increases particularly for LP none. This suggests maybe a few minorities in some former colonies may learn English as a first language

Finally, in column (6) we use information on parental country of origin available in the 2001 and 2006 censuses to include a set of interactions between second generation status (individuals born in Canada to at least one immigrant parent) and the linguistic proximity of the country of origin of the individual's father. If the LP variable is picking up other attributes that may affect labour force outcomes – such as occupational or cultural preferences or, most importantly, ethnicity – that are likely passed on to the next generation by first immigrant parents, these interaction terms may be relevant for wage determination. However, to the extent that LP is a good

[16] The estimated coefficient for genetic distance, once POB is controlled for, is positive which could indicate, as Isphording and Otten (2014) note in their paper, positive selection in unobserved ability or effort. It is negative when POB are excluded.

proxy for language proficiency, parental LP should not affect the earnings of second generation migrants compared to third generation (plus) Canadians. Results suggest that this indeed is the case. The coefficients for these interactions are small and not significant, both on their own (not shown here) and when the model also includes indicators for the father's broad area of origin (POB). If anything, second generation children of parents coming from English-speaking countries have a 1.7 per cent advantage over the third generation (plus). Overall, Table 3 results suggest that our LP measure is a good proxy for linguistic proficiency.

In a similar spirit, we analyze whether immigrants arriving from countries with languages closer to English hold "better" jobs than those with less linguistic proximity. We characterize good jobs as those that require relative high levels of social and analytical skills and low amounts of physical strength. A simple descriptive analysis (without any socio-demographic controls), shown in Table 2A in the appendix, offers a mixed picture, as immigrants with no linguistic proximity appear to hold better jobs than those with low levels of LP. Further, those arriving from English-speaking countries hold worse jobs than those with medium levels of LP.

To estimate how the characteristics of the jobs immigrants hold vary with LP, we use skill indexes as dependent variables in our basic regression (1). For the purposes of the exposition, we focus on Strength, Social and Analytical skills that better represent manual and cognitive skills. The indexes measure skills from the demand side of the market, i.e. the skills involved in performing each job, not the actual skills of the worker. They are also interpretable so that one unit change of the index indicates a one standard deviation change in the skill distribution of Canadian workers. After controlling for basic measures of human capital as well as POB, the first three columns of Table 4 show that immigrant men with the most distant languages typically hold jobs requiring more strength skills (0.15 standard deviations (SD) more) and fewer analytical or social skills (0.13 and 0.26 SD less respectively) than the native born. Immigrants from English-speaking origins, on the other hand, hold jobs requiring fewer strength skills (-0.07) and more analytical and social skills (0.09 and 0.06 respectively) than natives.

Similar to our analysis in Table 3, columns (4) to (6) in Table 4 include indicators of LP of immigrant' fathers' country of origin to estimate the extent to which LP might be picking up parental characteristics, other than language, passed on to the next generation. For instance, if an immigrant household with a more distant LP experienced hardship upon arrival, their (Canadian) children may have less advantageous upbringing that is reflected in poorer labour market outcomes. However, if LP only proxies linguistic ability of the individual, we should not observe any differences when using parental LP among second generation immigrants. Consistent with the findings for wages, we do not see a significant effect of father's LP on the level of strength skills employed by second generation men. Father's LP is not significantly correlated with the analytical and social skills required in their children's jobs, except for some weak association for low levels of father's LP and analytical skills.

3.2. The effect of linguistic proximity in the medium and long term

While estimates in Table 3 capture the average effect of LP on immigrants with different lengths of stay in the country, a relevant question is whether or not LP is related to the rate of immigrant assimilation over time in Canada. It is plausible that closer linguistic proximity allows for faster learning and that this translates into a more rapid progression and labour market success that widen the initial wage differences by LP. Conversely, any initial disadvantage due to linguistic distance could close over time as people become proficient in the local language and individuals arriving from countries with no connection to English slowly catch-up to those with closer linguistic proximity. To unveil what pattern better fits the data, we extend the basic model to consider a fully flexible specification of the arrival-cohort movements through the four census years in the sample. We follow immigrants with (potentially) different levels of linguistic proficiency by their entry cohort over time as in Borjas (1985) to examine the role of language fluency in labour market outcomes with time in Canada.

$$\log(w\quad) = X\beta_1 + \beta_2 \sum_l \sum_t \sum_j Coh_j * yr_t * L_l + \beta_3 \sum_t year_t + \epsilon \qquad (2)$$

where Coh_j are indicators designating j five-year immigrant arrival cohorts, that are interacted with the indicators for survey year (yr_t) and the five linguistic proximity indicators (L_l). Cohorts are defined over five years (i.e. the 86-90 cohort includes individuals arriving between 1986 and 1990) and followed since entry through the 1991-2006 Census years. Thus, Equation (2) includes 50 relevant interactions (10 *cohort*time* indicators times 5 indicators of linguistic proximity). Note that this is a more flexible approach than that using only a variable of years since migration (and its square), since the rates of change are not constrained to be the same across arrival cohorts. Therefore, β_2 is the vector of coefficients that tracks the assimilation of immigrant cohorts by linguistic proximity. The model also includes an indicator for all cohorts arriving before 1986, so that the coefficients are relative to the native-born.

Figure 2 shows estimated weekly wages by years in Canada for each arrival cohort of immigrant men relative to similar natives. For simplicity, the graph only includes results for immigrant men whose languages share no branches with English (denoted by *LP=None* and represented with the dotted line) and for those arriving from English-speaking countries (denoted by *LP=Same* and represented with the continuous line). Weekly wages for other levels of LP lie in between those lines. A full set of estimates for the wage assimilation of all linguistic groups can be found in the electronic appendix (Table 3A).

The weekly wages of those whose languages have no connection to the host country language are between 40 and 30 percent lower at arrival than those of the native born. Most arrival cohorts experience some wage improvement, around a 10% increase, during the first five to ten years in the country, but very little after that. Even after 15 years in Canada, their wages remain between 20 and 25 percent below those of Canadian born. The profile of wages of immigrants with the closest linguistic proximity is quite flat and their levels indistinguishable, for the most part, from those of the native born. Consistent with the above mentioned literature that points to declining fortunes for current immigrants to Canada, recent arrival cohorts (96-00) experience wider differences at arrival by LP than those who arrived in the late 1980s.

Finally, to assess labour market integration over time in Canada, we estimate a set of fully flexible models of the different skills required by jobs that include three-dimensional interactions of LP, arrival-cohort and time in Canada as in equation (2). If LP facilitates the acquisition of the local language, the job standing of immigrants with less distant languages, as measured by their use of "better job" skills (more social and analytical skills and less physical strength), should be higher, and/or improve faster over time than for other migrants.

Figure 3 displays how the skill requirements of the jobs that immigrants hold (relative to similar natives) change with years in Canada for different levels of LP to English. As before, we only show results for those with either no or complete proximity to English. Results for the other levels of linguistic proximity lie in between both cases and are available upon request.

Overall, there is little evidence that skill requirements improve (or change much) with years since arrival. Immigrants for whom English is the most used language in their country of origin (represented by the continuous line), work in jobs that are very similar to those of the native-born in terms of strength and social skills, and slightly better in terms of analytical skills. Further, their relative position either remains stable or converges toward natives over time. At arrival, immigrants with no linguistic proximity to English work in jobs requiring around 0.1 to over 0.2 SD (for the most recent arrival cohorts) of strength above those required to the average worker; between 0.3 to 0.4 SD lower social skills, and around 0.25 lower analytical skills. The levels of strength required in their jobs rise slightly with time in Canada and for successive cohorts, whereas social and analytical skills remain relatively stagnant (or even decline) over time for all cohorts. The gaps in skill requirements by LP are larger for the 96-00 cohort compared to the 86-90 cohort.

4. **The effect of LP on labour market outcomes of educated workers**

Given the emphasis of Canadian policy during the 1990s and early 2000s on selecting immigrants with high levels of human capital, in particular high levels of educational attainment, it is natural to question whether or not the transfer of skills embodied in formal education depends greatly on the linguistic ability of immigrants. To assess this, we introduce an interaction between tertiary education and LP in the basic model:

$$\log(w \quad) = X \quad \beta_1 + \beta_2 univ + \beta_3 \sum_l \; L_l \quad + \; \beta_4 \sum_l \; univ * L_l \quad + \epsilon \qquad (3)$$

where *univ* is an indicator for whether the individual has university education and all other variables are defined as in (1). β_2 estimates the returns to university education for the native born, the vector β_3 estimates the effect of different levels of LP for non-university educated immigrants (relative to non-college native born workers) and the vector β_4 estimates the joint effect of LP and university education for immigrants. Column (1) of panel A in Table 5 reports the coefficients of this wage regression, while panel B computes the total returns of university educated immigrants with varying degrees of LP, relative to the reference group of non-educated native born (by adding β_2 and the appropriate elements of β_3 and β_4 as indicated).

University-educated immigrants perform better than native-born workers with no tertiary education, but, except for those with *same language* or the few arriving from Nordic countries (2nd level), their earnings are lower than those of university-educated native-born. Wage penalties for low linguistic proximity (none or just one branch of the linguistic tree in common) are relatively larger for the university-educated (relative to university educated native-born) – around 26% (0.65 – 0.40) to 35% (0.65-0.31) lower wages – than for the non-educated (relative to similar natives) - only between 18% and 19% lower. There are significant wage differences for those with 3 levels of linguistic proximity, but these are lower for the university-educated than for the non-educated

(-0.13 versus -0.18). Finally, there are no significant wage differences between natives and immigrants arriving from English-speaking countries for any level of education.

Columns (2) to (4) in Table 5 display estimates of model (3) separately for each of the three skill indexes. Less-educated immigrants whose languages share at least two levels with English work in jobs requiring slightly higher (or similar) social skills, more analytical skills and lower strength skills than similarly-educated natives. By contrast, jobs held by low-educated migrants with either no or little connection to English are significantly inferior. The same gradient is apparent among the high educated. In the first row of panel B, university-educated immigrants, from English-speaking countries, seem to work in slightly "better" jobs than those held by similarly-educated native-born, as they require higher analytical (1.45-1.42 = 0.03), and social skills (1.64-1.54 = 0.10), and slightly lower strength skills (-1.36 – (-1.33) = -0.03). Educated immigrants with low LP, on the other hand, work in jobs requiring much lower levels of social and analytical skills and more strength skills than the educated native-born. Skill differences for those with low LP levels are larger among the university-educated than among less-educated migrants compared to their peers. For instance, the gap in terms of strength skills for high-educated arriving from countries sharing either no or only one linguistic branch with English is 0.27 and 0.26, respectively; whereas the same gap among the less educated is only 0.09 to 0.20. For social skills those numbers stand at -0.37 and -0.41 for high educated respectively, compared to -0.19 to -0.31 for the low-educated. A similar finding comparing high- and low-educated immigrant women to Canada with respect to similarly educated natives (for average LP) is reported in Adserà and Ferrer (2016).

The finding that LP does not affect the evolution of job skills with years since migration, and only affects the level at arrival, is surprising since expected improvements in fluency could help immigrants to move into better jobs over time. The result is particularly puzzling, considering the relatively high levels of education of Canadian immigrants. One would expect that education enables immigrants to gain job status even if they are initially under-placed. Indeed, it would seem that results in Fig. 3 may mask differential dynamics for university- versus non-university-educated immigrants. When we estimate skill progression for the sub-sample of university educated immigrants with respect to similar natives, results differ. These estimates, reported in Figure 4, show some improvement in the quality of jobs as migrants stay longer in Canada, as was expected. We observe a modest reduction in strength and a modest increase in social skills requirements with time in the country, even for those without any connection to English. Most importantly, there is a substantial increase in the analytical skills required by jobs for these immigrants.[17] We posit that even though those migrants may still have limited linguistic proficiency, they are able to update their credentials or access jobs that employ their formal education in sectors where language is not as relevant as in others that require more social interaction. Hence, both linguistic proximity and time in Canada somewhat affect the type of jobs educated immigrants hold over time.

5. A broader linguistic metric

Our results so far are based on a measure of the distance between the *most used* languages in both countries. To provide a more robust assessment of the role of linguistic ability, we use an additional indicator ("*Any official/major*") that measures the minimum distance between either English or French (the two official and most widely-spoken languages in Canada) and any of the official plus two most widely-spoken languages in the country of origin of the immigrant. This metric produces a more flexible and generous measure of linguistic proximity than the one based on the most used language as it encompasses any potential link between the languages spoken in the two countries. If immigrants are positively selected from the population in origin, they are more likely to be familiar with either English or French – even if none is their usual language – since, for example, they may have studied it in school as one of their country's official languages. Thus, to the extent that immigrants are positively selected, this more generous measure may more accurately reflect their linguistic ability.

There are some significant differences between the "*most-used*" and *"any official/major*" metrics. First, the former considers LP only to English, whereas the latter considers both Canadian official languages. Hence, their respective accuracy might be affected by the immigrant's settlement choice between provinces with or without large French-speaking populations.[18] Second, countries where English or French are either an official language or among the two most-used languages, such as France, India, Philippines, Kenya, Cameroon or Madagascar have the same level of LP under the "*any official/major"* metric as countries like the US or the UK, but they have

[17] Similar results are reported for all immigrant women – without controlling by LP – in Adsera and Ferrer (2016).
[18] An exploration of Canadian bilingualism and the settlement of immigrants across Canada is out of the scope of this paper, although partial results can be found in the working paper version of this article.

low LP ("none" or first level LP) when using the "*most-used*" metric instead. Third, under the "*any official/major"* metric, countries where romance languages are spoken (such as Italy, Portugal and all Spanish speaking countries) share four branches of the linguistic tree with Canadian official languages via their proximity to French. There are no languages at the four level of proximity under the "*most-used"* metric, when only English is considered, and romance languages share only one level of the linguistic tree with English since they are both Indo-European. Table 3A in the electronic appendix shows the main countries under this metric. The distribution of immigrants across linguistic groups shifts considerably compared to when most-used language is employed, and is more balanced among different levels of LP, with the exception of the second and third levels that continue to account for a tiny fraction of all immigrants. Immigrants classified as having no LP or sharing only one common branch with either official language account for 18 and 16 percent of the sample, while those in the new 4th level account for 18 percent. The largest group now includes immigrants classified as English (or French) speakers.

Even though the pattern of results is similar across metrics, the *any official/major* metric implies smaller differences in labour market outcomes across levels of linguistic proximity. Average weekly wages, in a model similar to column (3) in table 3, are only 16 percent lower than natives among those with no connection to English and 7 percent lower for those arriving from a country that has either English or French among its official or most spoken languages. The basic patterns of job skills are also similar, with immigrants holding jobs that require more strength and generally less analytical and social skill than the native born. Interestingly, the group holding the "worst" jobs when using *any official/major* metric are those sharing one branch in common with English or French, rather than those with no association. Immigrants with no LP (such as Chinese) access jobs with similar wages and high levels of social and analytical skills relative to those who speak the same language (a group that includes now Philippine and Indian immigrants). These estimates are available upon request.

Compared to the pattern displayed in Fig. 2, estimated wages of those with the closest proximity to either English or French increase with time in the country when *any official/major* is employed, but start 25 to 15 percent below those of natives. The fact that wage assimilation takes longer when we broaden our measure compared to when we use *most-used* language suggests that exposure to English or French at origin is imperfect among many who migrate from countries now classified as high LP. The measure likely overestimates initial proficiency while language acquisition takes place in Canada. Alternatively, it can signal other types of barriers, such as lack of networks, which migrants overcome with time in the country. Either explanation is supported by the fact that wage assimilation still takes place at a faster rate for those immigrants with potentially close exposure to English or French under the new metric - even if only imperfect exposure - than for those with no English or French among official/major languages in their country (Figure 1A in the appendix).

6. Discussion

In this paper we assess the role that language plays in the labour market performance of immigrants relative to the native born in a variety dimensions. In particular, we look at weekly earnings and at the occupational mobility of immigrants measured by the skill content of the jobs they hold at arrival and how this changes with years since migration. We combine the large samples in the restricted files of the Canadian Census (1991-2006) with a new measure of proximity of the most used language in the immigrant's country to that of the destination country (to proxy for linguistic proficiency) and with information on the occupational skills required by jobs immigrants hold.

Linguistic proximity affects the returns to human capital of immigrant men, with those arriving from countries whose languages have no or low connection to English displaying larger gaps with respect to natives. Though LP matters crucially in determining the level of returns, it does not seem to influence how fast (or whether) wage-parity is attained between immigrants and the native-born. However, we find that linguistic proximity influences the skills required in the jobs educated immigrants hold. We observe a modest reduction in strength, and increase in social skills, together with a substantial increase in the analytical skill requirements with time in the country, even for those without any connection to English. A larger linguistic distance imposes larger wage penalties to the university-educated than to those with lower educational attainment, relative to similarly-educated natives, and it significantly affects the status of the jobs they hold. To the extent that immigrants might be positively selected on the basis of their language ability, these results may underestimate the importance of linguistic proximity on the outcomes of immigrants and constitute a lower-bound of the penalty of poor linguistic ability. We feel confident that our measure of LP is indeed a good proxy for linguistic ability, rather than picking up other traits that could be associated with country of origin. In particular, LP does not seem to capture ethnicity, since the LP of first generation parents has no significant impact on the wages and skills employed by their sons.

The lack of wage convergence and indeed the small rate of job status mobility experienced over time by individuals with no linguistic proximity to Canadian languages are somewhat surprising. According to the human capital model, this could result, for example, from lack of improvement in language skills within *ethnic enclaves* that limit economic opportunities.[19] In separate estimates available upon request we find the size of linguistic community in province of residence to be negatively related to earnings. However, it is possible that there are other systematic barriers that hamper integration and that may or not be related to language proficiency. This could be the case if immigrants arriving as adults never reach a minimum level of proficiency required to access certain types of high-level jobs. Further, labour market institutions or migrant selection may also matter in the ability of migrants to become more fluent in the destination language. Isphording and Otten (2014) show that there is an increase in the proficiency gap by linguistic distance (measured with the Levenshtein index) with time in the country in the US whereas the reverse is true for Germany. Our findings for labour market outcome dynamics in Canada seem to point to a middle-ground where the gaps are never closed nor widen sharply with years since migration.

Migration studies are typically compelled to trade off the advantages of large samples that allow detailed analysis of immigrant characteristics with other essential information such as longitudinal records usually available in much smaller samples. Our study suffers from the same problem. Even though the census affords large samples, we cannot address whether the correlation between fluency and performance over time is due to unobserved innate characteristics of immigrants. As result, we can only interpret these findings – particularly those of changes over time – as associations between LP and the aggregated performance of immigrants. Nevertheless, our results agree with those of previous longitudinal studies in smaller samples that show language

[19] Adserà and Pytlikova (2015) show that the linguistic distance between a migrant's native language and the language of the destination country matters significantly less in migration decisions when there are large ethnic or linguistic enclaves in destination countries. However, migrants living and working within a relatively closed community may have difficulties in acquiring the destination language.

proficiency complements occupational skills (Berman et al. 2008) and affects the set of jobs accessible to immigrants at arrival (Imai et al. 2014).[20]

Overall our findings indicate an important role of linguistic ability in determining labour market performance (particularly for high-educated migrants). Promoting language training both in destination and in migrant-sending countries stands as a natural policy for immigrant integration that has returns not only for individuals but also for society as a whole by raising workers productivity and easing socioeconomic integration of new arrivals (Belot and Ederveen 2012, Chiswick and Miller 2015). These findings should be of particular interest to countries whose immigration policies aim to attracting high-skilled immigrants such as Australia and Canada, and for those contemplating such policies in the near future (OECD, 2014).

[20] Another concern when interpreting these findings is the possibility that participation in the labour market may change with years since migration in a way that is correlated with LP. We believe that this is a minor issue when analyzing the sample of migrant men since their level of labour force participation is high and stable.

References

Adserà and Ferrer, (2014) "The Myth of Immigrant Women as Secondary Workers: Evidence from Canada" *American Economic Review papers and proceedings*, 104(3), pp.360-364

Adserà and Ferrer, (2016) "Occupational Skills and Labour Market Progression of Married Immigrant Women in Canada", *Labour Economics* Vol 39, p. 88-98, April.

Adserà, A. and Pytliková, M. (2015) "The Role of Languages in shaping International Migration", *Economic Journal* Vol. 125, Issue 586, pp. F49-F81.

Antecol H, Kuhn P and Trejo S (2006) Assimilation via Prices or Quantities? Sources of Immigrant Earnings Growth in Australia, Canada and the United States. *Journal of Human Resources* 41: 821-840

Autor, D. (2013) "The "Task Approach" to Labor Markets: An Overview" NBER WP# 18711.

Autor, D. and D. Dorn (2013) The Growth of Low-Skill Service Jobs and the Polarization of the U.S. Labor Market, *American Economic Review*, 103(5), 1533-1597, 2013.

Autor, D. and M. J. Handel (2013). "Putting Tasks to the Test: Human Capital, Job Tasks, and Wages," *Journal of Labor Economics*, vol. 31(S1), pp. S59 - S96.

Autor, D., F. Levy, and R. Murnane. (2003). "The Skill Content of Recent Technological Change: An Empirical Investigation." *Quarterly Journal of Economics*,118(3): 1279–1333.

Aydemir, A. and Skuterud, M. (2005) "Explaining the Deteriorating Entry Earnings of Canadian Immigrant Cohorts: 1966-2000", *Canadian Journal of Economics*, Vol. 38(2), pp. 641-72.

Bakker, D., Müller, A. Velupillai,V., Wichmann, S. Brown, C.H., Brown, P., Egorov, D., Mailhammer, R., Grant, A. and Holman, E.W. (2009). "Adding typology to lexicostatistics: a combined approach to language Classification." *Linguistic Typology*, vol. 13(1), pp. 169-181.

Begin, K., C. Goyette and W. C. Riddell (2010). "Revising Canada's Immigrant Selection System: An Assessment of Recent Changes" Background paper prepared for Citizenship and Immigration Canada.

Belot, M., and S. Ederveen. (2012) "Cultural and institutional barriers in migration between OECD

countries." *Journal of Population Economics* 25:3: 1077–1105.

Berman, E., K. Lang, E. Siniver (2003) "Language-skill complementarity: returns to immigrant language acquisition", *Labour Economics* 10 (1): 265-90

Borjas,G. (2013) "The Slowdown in the Economic Assimilation of Immigrants: Aging and Cohort Effects Revisited Again" , *NBER Working Papers* 19116,

Borjas, G. (1985) "Assimilation and Changes in Cohort Quality Revisited: What Happened to Immigrant Earnings in the 1980's?" *Journal of Labor Economics* 3, pp. 463-89.

Bratsberg, B. and O. Raaum (2004) Identifying Earnings Assimilation of Immigrants under Changing Macroeconomic Conditions. *Scandinavian Journal of Economics* 106: 1-22

Cavalli-Sforza, L. L., Menozzi, P. and Piazza, A. (1994). *The History and Geography of Human Genes*. Princeton, NJ: Princeton University Press.

Chiswick, B. R. (1986) "Ìs the New Immigration Less Skilled Than the Old?" *Journal of Labor Economics*, 4(2):168-192

________(1978) “The Effect of Americanization on the Earnings of Foreign-Born Men”. *Journal of Political Economy* 86(5): 897-922.

Chiswick, B. and P. Miller (2015). “International migration and the economics of language.” In: Chiswick, B. R., and P. W. Miller (eds). *Handbook on the Economics of International Migration.* Amsterdam: Elsevier, pp. 211–270.

______________ (2010) “Occupational Language Requirements and the Value of English in the US Labor Market.” *Journal of Population Economics* 23(1): 353–372.

________(2007) “Computer Usage, Destination Language Proficiency and the Earnings of Natives and Immigrants.” *Review of the Economics of the Household* 5 (2): 129-157.

_________ (2003) “The complementarity of language and other human capital: immigrant earnings in Canada” *Economics of Education Review* Vol. 22. No.5. p. 469–80.

_________ (2002) “Immigrant Earnings: Language Skills, Linguistic Concentrations and the Business Cycle.” *Journal of Population Economics* 15(1): 31-57.

Dustmann, C. (1994) “Speaking Fluency, Writing Fluency and Earnings of Migrants.” *Journal of Population Economics* 7: 133–56.

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

----------------- (2002) “Language and the Earnings of Immigrants.” *Industrial and Labor Relations Review* 55 (3): 473–492.

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

Ferrer, A., G. Picott and C. Riddell (2014), “Canada’s Immigration Policy: Recent changes and developments” *International Migration Review,* vol. 48(3), pp.846-867

Ferrer, A.M., and C. Riddell (2008), “Education, credentials and immigrant earnings”, *Canadian Journal of Economics*, vol. 4(1), pp. 186-216

Ferrer, A.M., D. Green, C. Riddell (2006), “The Effect of Literacy on Immigrant Earnings”, *The Journal of Human Resources* vol. 41(2) p. 380-410

Friedberg R (2000) “You Can’t Take it With You: Immigrant Assimilation and the Portability of Human Capital. *Journal of Labor Economics* 18: 221-51

Imai, S. & D. Stacey & C. Warman (2014). *"From Engineer to Taxi Driver? Language Proficiency and the Occupational Skills of Immigrants*," Working Paper Series 18, Economics Discipline Group, UTS Business School, University of Technology, Sydney.

Immigrant and Refugee Protection Act, 2002 (http://laws.justice.gc.ca/eng/acts/i-2.5/page-1.html#h-2)

Isphording, Ingo E. and Sebastian Otten (2014) Linguistic Barriers in the Destination Language Acquisition of Immigrants, *Journal of Economic Behavior and Organization,* 2014, 105, 30–50.

Kossoudji, S. (1988). “The Impact of English Language Ability on the Labor Market Opportunities of Asian and Hispanic Immigrant Men.” *Journal of Labor Economics* 6 (3): 205-228.

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

Li, Qing and Arthur Sweetman, 2014. "The Quality of Immigrant Source Country Educational Outcomes: Do they Matter in the Receiving Country?", *Labour Economics* 26 pp. 81-93.

Oreopolous, P. (2011) "Why Do Skilled Immigrants Struggle in the Labor Market? A Field Experiment with Thirteen Thousand Resumes", *American Economic Journal: Economic Policy* 3 pp. 148–171

OECD (2014). "Who should be admitted as a labour migrant?", *Migration Policy Debates* 4, http://www.oecd.org/els/mig/OECD%20Migration%20Policy%20Debates%20Numero%202.pdf

Schaafsma, J. and A Sweetman (2001). "Immigrant earnings: age at immigration matters", *Canadian journal of Economics*, vol. 34(4), pp. 1066-1099

Spolaore, E. and Wacziarg, R. (2009). "The diffusion of development", *Quarterly Journal of Economics*, vol.124 (2), pp. 469-530.

Warman, C. and C. Worswick (2015) "Technological Change, Occupational Tasks and Declining Immigrant Outcomes: Implications for Earnings and Income Inequality in Canada," *Canadian Journal of Economics,* vol. 48(2), pages 736-772

Table 1. Sample Summary Statistics. Census 1991-2006

		Canadian born	Immigrant
Age		33	43
Marital Status		0.56	0.82
Education	HS or less	0.49	0.37
	Post-secondary (other)	0.35	0.31
	Bachelor	0.13	0.18
	Graduate	0.03	0.12
Labour Force Participation		0.95	0.96
Job skill index [1]	Social	-0.13	-0.09
	Strength	0.32	0.15
	Analytical	0.01	0.13
			Most used
Linguistic proximity to English [2]	None		0.48
	1st Level		0.36
	2nd Level		0.00
	3rd Level		0.03
	Same Language	1.00	0.13
Levenshtein [3]			82.66
# Observations		802,590	585,783

(1) The skill index is calibrated and normalized to the Canadian native born population (men and women) in 2001, so that 0 corresponds to the average Canadian worker in 2001.
(2) Linguistic Proximity for the "Major" metric it corresponds to of the highest common branch in the linguistic tree between the most used language at origin and English (most commonly used language in Canada)
(3) Levenshtein measures linguistic distance from 0 (same language) to around 106.

Table 2. Main immigrant source countries in Canada by Linguistic Proximity (LP) to English

	Most used language (%) [1]				
					Row Total
None	China [2] (26%) Philippines (10%)	Vietnam (8%)	Jamaica (5%) [3] Pakistan (4%)	Other (47%)	100%
1st Level	India (21%) Italy (13%)	Poland (9%) France (9%)	Portugal (8%) Sri Lanka (6%)	Other (34%)	100%
2nd Level	Denmark (60%) Sweden (20%)			Other (20%)	100%
3rd Level	Germany (46%) Netherlands (26%)			Other (28%)	100%
Same Language	US (21%) UK (62%)	Trinidad (10%)		Other (8%)	100%

(1) Percentages correspond to the shares among all countries in the corresponding LP category
(2) Includes the republic administrative region of Macao. Macao's most used language is either Cantonese or Mandarin. Only a small fraction of the population speaks Portuguese, the other official language
(3) Jamaica and Guyana's official language is English, although the most used language is an English based-Creole language, which is the reason they are classified as having LP = 0.

Table 3. Weekly wages of Immigrant men relative to natives

	First Generation and LP					Second Generation and LP
	(1)	(2)	(3)	(4)	(5)	(6)
1st Gen Immigrant		-0.04**				-0.34**
1st Gen*LP indicators:						
None	-0.38**		-0.21**	-0.22**	-0.25**	
1st Level	-0.29**		-0.24**	-0.24**	-0.25 **	
2nd Level	-0.06*		-0.02	-0.02	-0.03	
3rd Level	-0.22**		-0.18**	-0.18**	-0.19**	
Same Language	-0.05**		-0.02**	-0.02**	-0.12**	
Levenshtein Index		-0.003**				
Mother Tongue=English					0.10**	
2nd Gen *LP indicators:						
2nd Gen*Father No LP						-0.06
2nd Gen*Father 1st Level						0.01
2nd Gen*Father 2nd Level						-0.004
2nd Gen*Father 3rd Level						0.029
2nd Gen*Father Same						0.017*
Can. Father–Imm Mother						-0.02**
POB control [1]	No	No	Immig.	Immig.	Immig.	Father's
Genetic distance	No	No	No	Yes	No	No
Observations	1,387,565	1,387,565	1,387,565	1,353,565	1,387,565	914,245

All regressions in first three columns include, controls for marital status, experience, experience squared, education, Canadian location, GDP per capita at country of origin at time of migration and survey year. Individuals are 18 to 60 years old. Immigrants are adults at arrival. The reference group in columns (1)-(5) are all natives and in columns (6) are 3rd generation plus natives. Data from Census 1991-2006 in columns (1)-(5) and 2001-2006 in column (6).

LP measures proximity between the most commonly used language at origin and the most commonly used language at destination (English). The Levenshtein index measures linguistic distance to English (0 to 106). Genetic distance measures distance between the largest ethnic groups in the two countries.

(**, *) indicates the coefficient is significant at 1% or 5% respectively

[1] The omitted POB (place of birth) group is US/Europe

Table 4. Skills required in jobs of immigrant men relative to natives, by generation and LP

	First Generation and LP			**Second Generation and LP**		
	(1)	(2)	(3)	(4)	(5)	(6)
	Strength	Analytical	Social	Strength	Analytical	Social
1st Gen Immigrant				0.164**	-0.242**	-0.376**
1st Gen*LP indicators:						
None	0.15**	-0.13**	-0.26**			
1st Level	0.22**	-0.23**	-0.35**			
2nd Level	0.01*	0.05**	-0.07**			
3rd Level	0.01	-0.01	-0.10**			
Same Language	-0.07**	0.09**	0.06**			
2nd Gen *LP indicators:						
2nd Gen*Father No LP				0.004	-0.102*	-0.061
2nd Gen*Father 1st Level				-0.005	-0.07*	-0.038
2nd Gen*Father 2nd Level				-0.000	-0.03	-0.023
2nd Gen*Father 3rd Level				0.025	-0.043	-0.033
2nd Gen*Father Same				-0.070**	0.034**	0.056**
Can. Father – Immig Mother				-0.066**	0.035**	0.065**
POB control [1]	Immigrant	Immigrant	Immigrant	Father's	Father's	Father's
# Observations	1,353,560			891,810		

All regressions include, in addition, controls for marital status, experience, experience squared, education, Canadian location, GDP per capita at country of origin at time of migration and survey year. Individuals are 18 to 60 years old. Immigrants are adults at arrival. The reference group in columns (1)-(3) are all natives and in columns (4)-(6) are 3rd generation plus natives. Data from Census 1991-2006 in columns (1)-(3) and 2001-2006 in columns (4)-(6).

LP measures proximity between the most commonly used language at origin and the most commonly used language at destination (English).

(**, *) indicates the coefficient is significant at 1% or 5% respectively

[1] The omitted POB (place of birth) group is US/Europe

Table 5. Effect of linguistic proximity of the most used language in origin to English and education on wages and required job skills of immigrant men
(relative to non-university educated native born)

		Most used Language			
		Wages	**Skills in jobs**		
			Strength	**Social**	**Analytical**
PANEL A. Model (3) estimates relative to non-university educated native born					
1	Non-University (NB)	Ref.	Ref.	Ref	Ref
2	Imm - Same Language	0.00	-0.12**	0.11 **	0.17**
3	Imm - 3rd Level	-0.18**	-0.04**	-0.04**	0.05**
4	Imm - 2nd Level	-0.02	0.00	0.01	0.10**
5	Imm - 1st Level	-0.19**	0.20**	-0.31**	-0.22**
6	Imm - None	-0.18**	0.09**	-0.19**	-0.10**
7	University (NB)	0.65**	-1.33**	1.55**	1.42**
8	Uni *x* Imm - Same Language	-0.01	0.10**	-0.01	-0.04**
9	Uni *x* Imm - 3rd Level	0.06**	0.13**	-0.07**	0.03**
10	Uni *x* Imm - 2nd Level	0.08	-0.07	-0.06	-0.03
11	Uni *x* Imm - 1st Level	-0.15**	0.06**	-0.09**	-0.10**
12	Uni *x* Imm - None	-0.07**	0.18*	-0.18*	-0.14*
PANEL B. University educated immigrant's total returns by LP					
Same Language (2) + (7) + (8)		0.64**	-1.36**	1.64**	1.45**
3rd Level (3) + (7) + (9)		0.53**	-1.25**	1.43**	1.37*
2nd Level (4) + (7) + (10)		0.71**	-1.40**	1.49**	1.49
1st Level (5) + (7) + (11)		0.31**	-1.08**	1.14**	1.23**
None (6) + (7) + (12)		0.40**	-1.06**	1.17**	1.29**

Columns in panel A show estimates of the effect of linguistic proximity and university education on wages and job-required levels of strength, social and analytical skills, relative to those of similarly educated native born, according to model (3)

Panel B computes *the total returns* of university educated immigrants with varying degrees of LP, relative to the reference group of non-educated native born.

All regressions include controls for marital status, experience, experience squared, location, GDP per capita at country of origin at time of migration, place of origin (US/Europe is the omitted group) and survey year. Individuals are 18 to 60 years old. Immigrants are adults at arrival. Observations are 1,387,565 for wage regression and 1,353,560 for skill regressions

(**, *) indicates the coefficient is significantly different from 0 at 1 and 5% respectively

Figure 1. Distribution of job-required skills by immigrant status

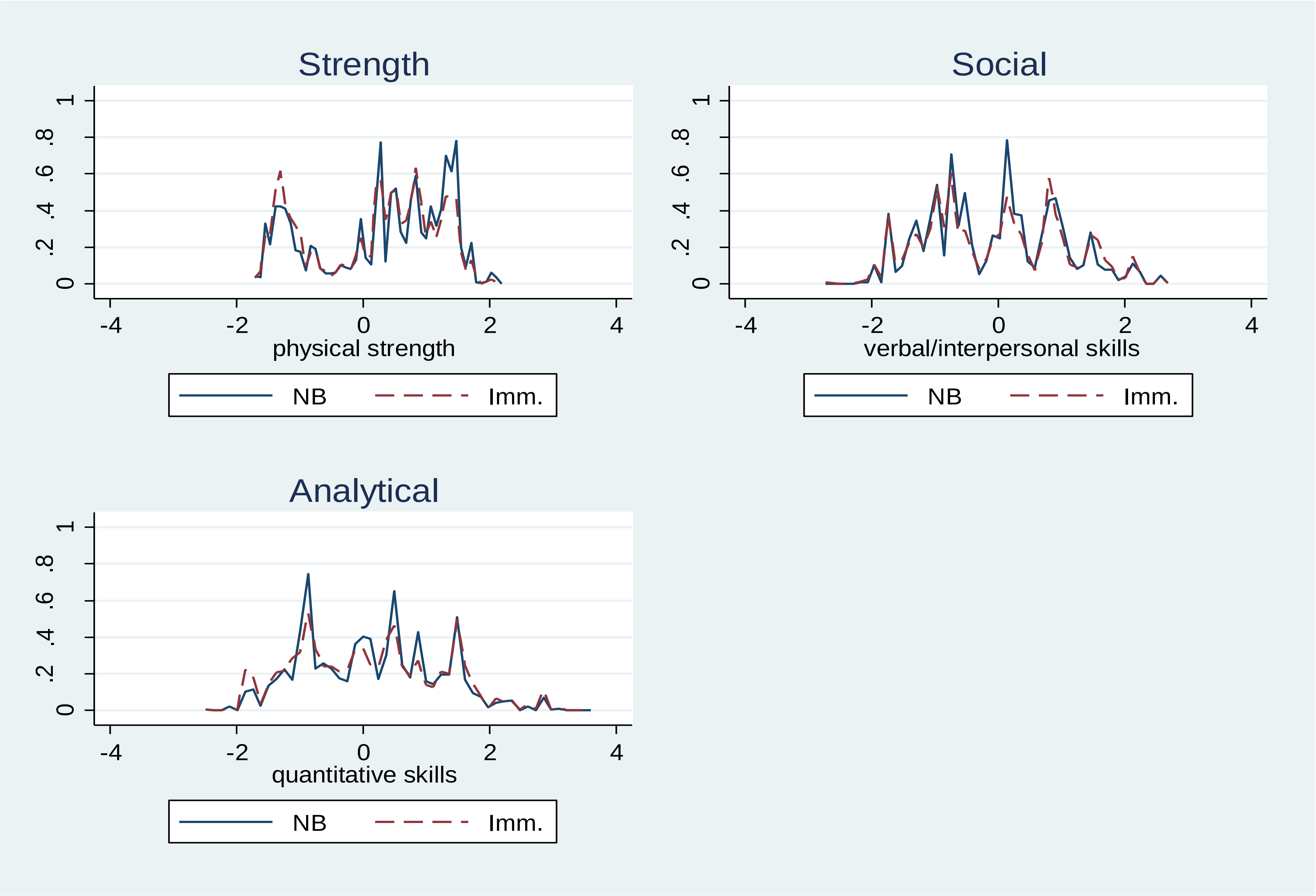
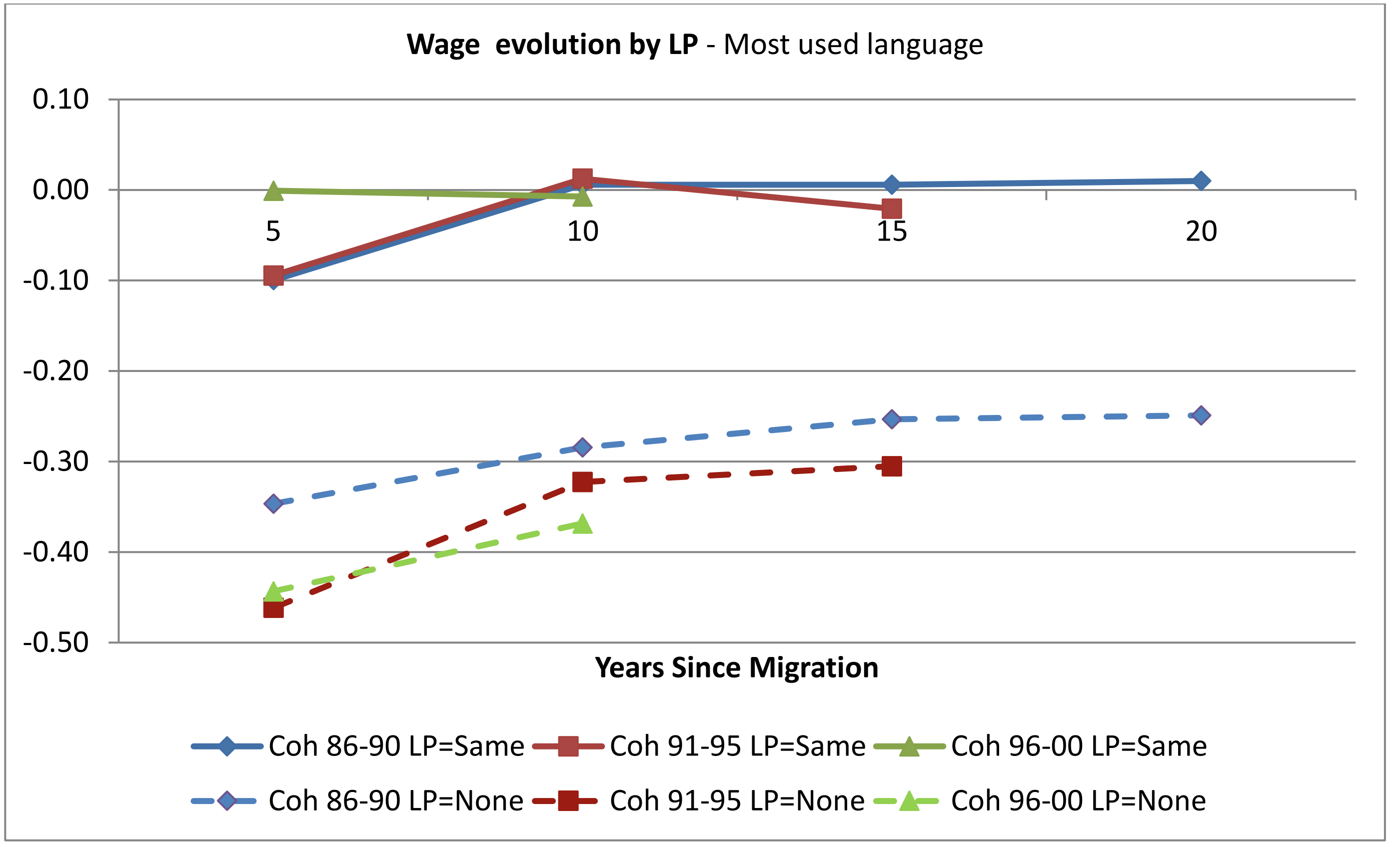


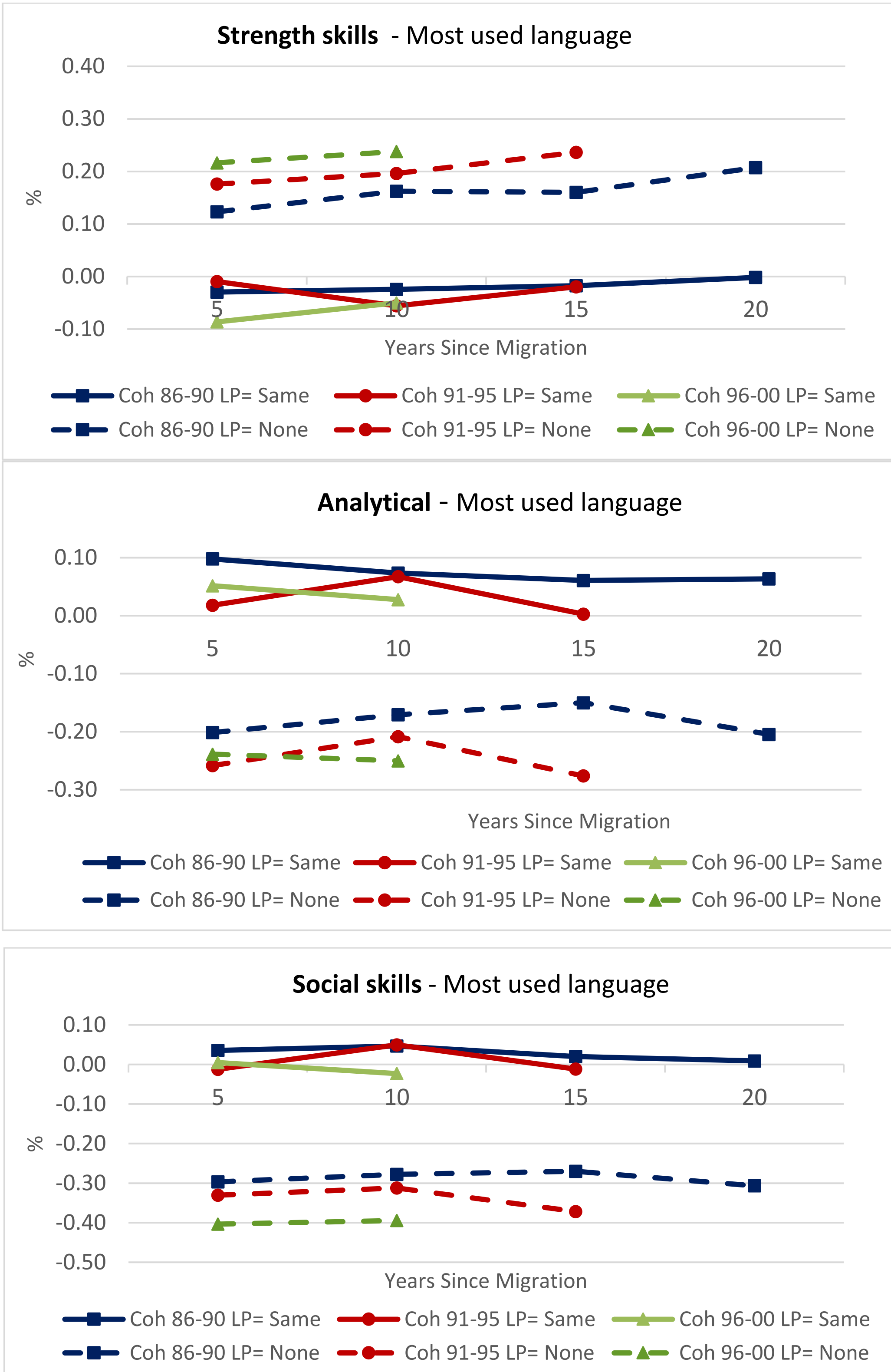
Figure 2. Immigrants' wage assimilation by Linguistic Proximity to *Most used* language and arrival cohort



Note: Model also includes marital status, experience, experience squared, education, location, area of origin indicators, GDP per capita of country of origin and survey year `

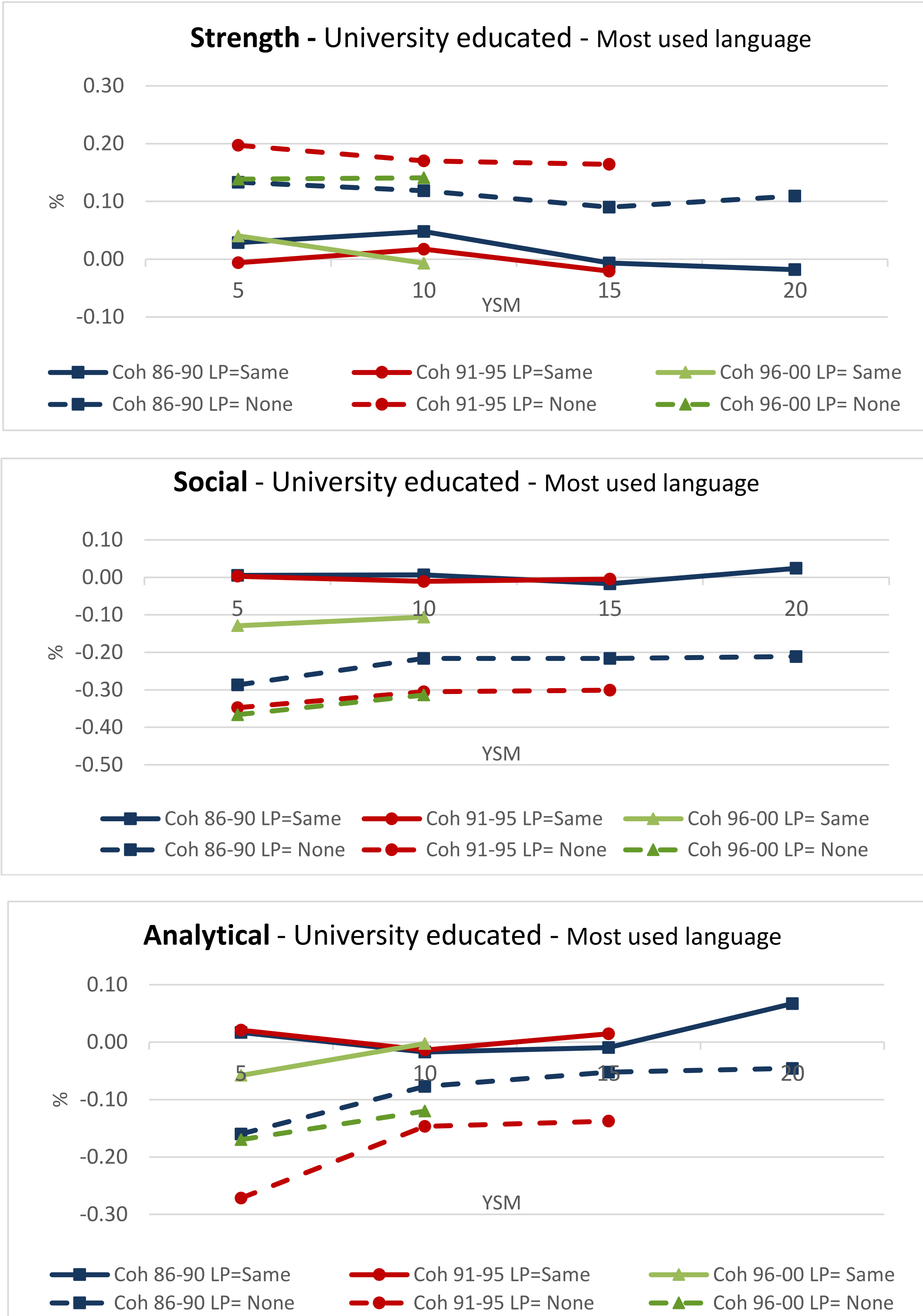
LP = None indicates languages at origin and destination do not share any branch of linguistic tree. *LP = Same* indicates both languages are the same or have at least the 5th branch of the linguistic tree in common.

Figure 3. Evolution of job-required skills by Linguistic Proximity and arrival cohort



See note to figure 2

Figure 4. Evolution of job-required skills by Linguistic Proximity and arrival cohort, University-educated men.



See note to figure 2

Appendix

Table 1A. Correlation of Skills employed in Jobs					
	Analytical	Social	Visual	Motor	Strength
Analytical	1.00				
Social	0.88	1.00			
Visual	-0.30	-0.38	1.00		
Motor	-0.49	-0.67	0.78	1.00	
Strength	-0.65	-0.75	0.67	0.89	1.00

Table 2A. Average skills required in jobs held by natives and immigrants by linguistic proximity			
	Social	**Analytical**	**Strength**
Native born	**-0.09**	**0.04**	**0.29**
Immigrant	**-0.09**	**0.12**	**0.16**
LP Most used language			
None	-0.08	0.13	0.08
1st Level	-0.32	-0.06	0.36
2nd Level	0.16	0.39	0.13
3rd Level	0.08	0.30	0.17
Same Language	-0.08	0.05	0.29

Table 3A. Coefficients of wage assimilation by cohort and Linguistic proximity

		LP Same		LP 3 branches		LP 2 branches		LP 1 branch		LP None	
Cohort	YSM	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Coh 86-90	5	-0.100	0.000	-0.243	0.000	0.057	0.030	-0.374	0.000	-0.347	0.000
	10	0.006	0.734	-0.169	0.000	0.015	0.050	-0.335	0.000	-0.284	0.000
	15	0.006	0.760	-0.146	0.000	0.204	0.000	-0.299	0.000	-0.253	0.000
	20	0.010	0.614	-0.150	0.000	-0.096	0.000	-0.258	0.000	-0.249	0.000
Coh 91-95	5	-0.095	0.000	-0.342	0.000	-0.101	0.000	-0.476	0.000	-0.462	0.000
	10	0.012	0.476	-0.174	0.000	0.205	0.000	-0.336	0.000	-0.323	0.000
	15	-0.021	0.305	-0.188	0.000	0.235	0.000	-0.339	0.000	-0.305	0.000
Coh 96-00	5	-0.001	0.971	-0.243	0.000	-0.110	0.000	-0.439	0.000	-0.443	0.000
	10	-0.007	0.761	-0.188	0.000	0.101	0.000	-0.362	0.000	-0.368	0.000
Coh 06	5	-0.124	0.000	-0.258	0.000	0.050	0.041	-0.521	0.000	-0.581	0.000

The regression includes, in addition, controls for marital status, experience, experience squared, education, Canadian location, GDP per capita at country of origin at time of migration, indicators for broad area of place of birth and survey year. Individuals are 18 to 60 years old. Immigrants are adults at arrival. The omitted area of origin group is US/Europe. The reference group are the native born.

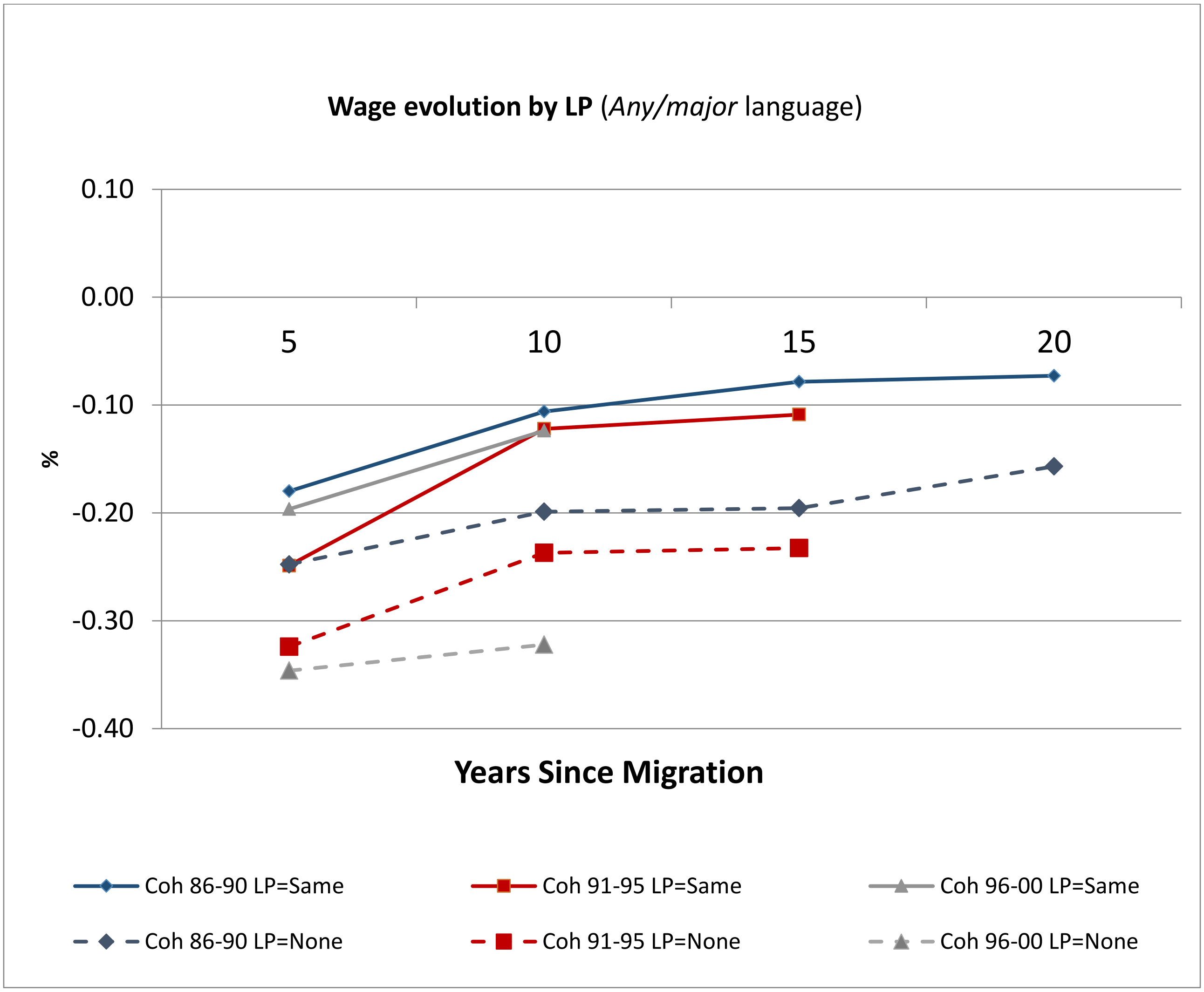
The figures in Column "LP Same" and "LP None" are represented graphically in Figure 2 in the text.

LP measures proximity between the most commonly used language at origin and the most commonly used language at destination (English).

Table 4A. Main immigrant source countries in Canada by Any official/Major language metric

LP from any official language at origin to English or French			
			% of all immigrants
None	China (37%), Vietnam (19%)	Other (44%)	0.19
1st Level	Poland (20%), Sri Lanka (13%), Iran (10%), Yugoslavia (6%)	Other (51%)	0.16
2nd Level	Denmark (60%) Sweeden (20%)	Other (20%)	0.00
3rd Level	Germany (42%), Netherlands (25%), Romania (25%)	Other (8%)	0.03
4th Level	Italy (27%), Portugal (18%)	Other (55%)	0.15
Same Language	UK (20%), India(18%), Philippines (10%), US (6%)	Other (46%)	0.47

Figure 1A. Immigrants' wage assimilation by Linguistic Proximity to *Any/Major* Language and arrival cohort



Note: Linguistic Proximity for the "Any/Major" metric corresponds to the highest common branch in the linguistic tree between any official and the two most widely spoken languages at origin and English or French (official languages in Canada).