



Effects of Linguistic Distance on Cognitive Skills, Health, and Social Outcomes in Canadian Immigrants

Victor Kuperman*

Department of Linguistics and Languages, McMaster University, Hamilton, ON, Canada

Learning languages is more difficult if they are more distant (dissimilar) from the person's first language. For immigrants, a greater distance between their L1 and official languages of their host country impedes integration. Using microdata from the PIAAC survey in 2012, we quantified the cognitive, and socio-economic impact of linguistic distance on a sample of 2,018 immigrants to Canada representing 12 diverse languages. Speakers of languages more distant from English or French showed lower literacy scores, worse health and weaker community engagement, yet numeracy and income were unaffected. We discuss causes and implications of these findings for immigration experience.

Keywords: linguistic difference, Canada, acculturation, health, immigration

OPEN ACCESS

Edited by:

Stephanie J. Nawyn,
Michigan State University, United States

Reviewed by:

Job Schepens,
Technical University Dortmund,
Germany

Jane Freedman,
Université Paris 8, France

*Correspondence:

Victor Kuperman
vickup@mcmaster.ca

Specialty section:

This article was submitted to
Refugees and Conflict,
a section of the journal
Frontiers in Political Science

Received: 11 February 2022

Accepted: 23 June 2022

Published: 15 July 2022

Citation:

Kuperman V (2022) Effects of Linguistic Distance on Cognitive Skills, Health, and Social Outcomes in Canadian Immigrants. *Front. Polit. Sci.* 4:874195. doi: 10.3389/fpos.2022.874195

1. INTRODUCTION

Proficiency in the country's official or majority languages determines to a large degree how well immigrants integrate into the host country (see among others Chiswick and Miller, 2005; Adserà and Pytliková, 2016). Often, these languages—labeled here La (or additional languages)—are different from the first language of the immigrant—labeled here L1—and require an extended learning period. Lower proficiency in La poses manifest barriers to the economic, professional, and academic success of immigrants, as well as to their acculturation and civic engagement (reviewed by Ginsburgh and Weber, 2020). Moreover, economic and cultural repercussions of the parental deficit in La proficiency propagate across generations (Bleakley and Chin, 2008; Casey and Dustmann, 2008). This is hardly surprising since literacy and other information-processing skills contingent on language are a central component of human capital, critical for functioning in the modern-day technological society (Chiswick and Miller, 1998; Bynner, 2004; Finnie and Meng, 2007). The present study focuses on a systematic factor that predicts group differences in the difficulty of acquiring proficiency in La, i.e., the linguistic distance between the immigrant's L1 and La.

The notion of linguistic distance captures the intuition that speakers of language A need less effort to become proficient in language B if these languages are relatively similar to each other, e.g., the languages share a larger share of their vocabulary, have identical or related phonological, morphological or syntactic features, and/or share a writing system. Speakers of language C that is more distant from language B in some or all of these linguistic aspects will require more effort and may ultimately achieve lower proficiency in B than speakers of A. For instance, with the same effort, an average speaker of a Romance language, Italian, may reach a higher proficiency in another Romance language, French, as compared to an average speaker of an Altaic language, Korean, learning French. In the model by Chiswick and Miller (1995), the group factor of linguistic distance between L1 and La, along with individual factors like education and talent, determine the learner's capacity to achieve high proficiency in a foreign language in a given time period and given set of incentives.

The intuitions about systematic differences in language learnability driven by differences between languages find a robust confirmation in research on second language acquisition (see among others Kellerman, 1979; Ringbom, 2006; Jarvis and Pavlenko, 2008). Perhaps the most compelling illustration of the role of L1-La linguistic distance on La proficiency comes from recent high-powered studies by Van der Slik (2010), Schepens et al. (2013b, 2016, 2020), and Van der Slik et al. (2015, 2019), see also Van der Slik (2010). All of these studies analyze scores from the speaking portion of the state exam of Dutch as a Second Language from over 50,000 speakers of 70 different language backgrounds (e.g., Schepens et al., 2020). Across multiple studies, Schepens et al. determined a unique negative effect of linguistic distance between the speaker's L1 and Dutch on speaking proficiency in Dutch, over and above a range of demographic variables related to the speaker (e.g., age, gender, education, other languages spoken) and their country of origin (national GDP, schooling, and others). L1 speakers of languages more similar to Dutch (e.g., German, English) obtained higher scores than L1 speakers of the languages lexically, phonologically, and morphologically remote from Dutch (e.g., Arabic, Turkish).

Since as early as the 1960's (Marschak, 1965), social science research has revealed that the impact of linguistic distance far exceeds the realm of academic assessments (see reviews by Borjas and Chiswick, 2019; Ginsburgh and Weber, 2020). While differing widely in their operationalization of linguistic distance, these studies converged on the economic relevance of linguistic distance. Immigrant speakers of languages more distant from the official language of the country (e.g., English in the USA, Hebrew in Israel) demonstrated poorer economic outcomes, including lower levels of employment and income (Beenstock et al., 2001; Chiswick and Miller, 2012; Ispording, 2014; Adserà and Pytliková, 2016; Clarke and Ispording, 2017). Beyond the economic sphere, a greater linguistic distance between L1 and La was shown to constrain one's ability or willingness for civic engagement with their local institutions or communities, e.g., volunteering in community organizations, voting, or working toward community improvement (Gottlieb and Gillespie, 2008; Barrett and Brunton-Smith, 2014; Adserà and Pytliková, 2016). An additional demonstrated disadvantage of speaking a more distant L1, compared to the host country's language, is seen in poorer health outcomes (Clarke and Ispording, 2017), which might be associated with more limited health literacy and willingness or ability to access health services because of linguistic barriers.

The present study aims to deepen the current understanding of linguistic distance between the immigrant's first language and the host country's language as a relevant factor for adaptation and acculturation of immigrants across multiple spheres of life. The study departs from previous work in two ways. First, instead of concentrating on a single outcome, we simultaneously consider a broad range of outcomes potentially affected by linguistic distance. These include proficiency in information-processing skills (literacy and numeracy), civic engagement (volunteering frequency and political efficacy), economic factors (income), and health.

Second, we concentrate on immigration to Canada, which shares some demographic and linguistic characteristics with comparable population groups in developed countries but differs in several other regards relevant to our research goals (see also Chiswick and Miller, 2003; Adserà and Ferrer, 2015). The majority (58% in 2016) of Canadian are economic immigrants "selected for their ability to contribute to Canada's economy" (Statistics Canada, 2016). To be considered for admission as an economic immigrant, the principal applicant must demonstrate a minimum of a higher-intermediate proficiency level in speaking, listening, reading, and writing in either English or French (level 7 or higher in the Canadian Language Benchmark test or Niveaux de compétence linguistique canadiens). Bonus points can be earned if either the principal applicant or their spouse demonstrates a lower-intermediate (level 5) or higher level of proficiency in a second official language (Immigration Refugees Citizenship Services, 2021). These and other selection criteria for the immigrant body suggest relatively strong mastery of at least one official language before immigration to Canada, and this mastery likely evolves further through immersion in the English- or French-speaking environment upon immigration to Canada. For a comparative literacy and numeracy advantage over several other Western countries in the 1 and 1.5-generation of Canadian immigrants (see Levels et al., 2017). While an average Canadian immigrant lagged behind Canadian-born counterparts in literacy and numeracy skill (Xu et al., 2017), the difference was minimal for economic immigrants and for holders of advanced degrees. It is thus, possible that the effects of linguistic distance reported in prior literature will be attenuated because of a relatively high range of La proficiency among immigrants to Canada.

Other mandatory criteria also include secondary or post-secondary education and work experience in designated priority areas, while pre-arranged jobs and Canadian education or work experience grant additional bonus points. As a result of these immigration policies, a larger percentage of immigrants than the Canadian-born population hold a bachelor's degree or higher (4 out 10 vs. one-quarter among 16–64 y.o. in 2016, <https://www150.statcan.gc.ca/n1/daily-quotidien/171129/dq171129a-eng.htm>); the advantage in the educational attainment persists for Canadian immigrants admitted as children vs the respective Canadian-born population. While the wage gap exists between Canadian immigrants and non-immigrants, there is evidence that it has been closing (Crossman et al., 2021). The selection based on human capital applied for most popular immigration types may diminish the impact of linguistic distance on economic outcomes, as compared to reports on immigrants from other countries.

Another feature of economic immigration in Canada is its skewness toward younger adults. The points system that is used to assess potential candidates for immigration grants a higher number of points to younger individuals, with the maximum of 100 points given to 20–29 year olds and 0 points to 45+ y.o. The preference for younger individuals has consequences for language learning as well, given the well-described effect of age on rate of foreign language acquisition (Muñoz and Muñoz, 2006). This feature, again, may undermine the base rate differences stemming from different L1s and their distance from

La. In sum, Canada presents an interesting test case to verify whether linguistic distance has an impact on a population group that has a more restricted positively-skewed range both in La proficiency, age, and in economic standing (Mayda, 2010; Belot and Ederveen, 2012). Arguably, if the effect of linguistic distance is observed in this population, it is likely to be observed more strongly in the non-economic immigration groups and in the countries where economic immigration accounts for a lesser fraction of newcomers.

We pursue our goal by harvesting rich data presented in the Survey of Adult Skills collected in 2011–2012 in Canada under the Programme for the International Assessment of Adult Competencies (PIAAC) guided by the Organization for Economic Cooperation and Economic Development (OECD). The PIAAC data includes an assessment of information-processing skills (literacy, numeracy, and problem solving) that are foundational for one's ability to succeed in the technological economies of the present day (PIAAC Literacy Expert Group, 2009; Programme for International Assessment of Adult Competencies, 2012). It also includes an extensive demographic questionnaire and a module querying how the skills are used at home and in the workplace. The nationally representative PIAAC Canada sample contains over 27,000 adult Canadians (16–65 y.o.) and it oversampled several populations, including immigrants, to provide detailed information on those groups. PIAAC data are supplied with weights, which enable generalization of findings over the entire Canadian population. The critical variable of all analyses below was linguistic distance between the participant's L1 and the test language, i.e., the La in which they chose to complete the PIAAC survey and assessments.

In sum, this study estimates the effect of linguistic distance on a broad range of outcomes related to skills, economic success, social integration, and health in Canadian immigrants, while controlling for multiple individual-level and country-level covariates. If Canadian immigrants align with those from other countries (see Section Discussion above), we anticipate that linguistic distance would constrain learnability of and proficiency in La. An implicit assumption in the prior literature is that the impact of linguistic distance on the economic or health outcomes of an individual is indirect and is (fully or partly) mediated by La proficiency. We test this assumption by estimating the amount of variance explained by linguistic distance in a range of dependent variables, while also controlling by La proficiency. To reiterate, our goal is to determine in which spheres of life Canadian immigrants from linguistically diverse backgrounds experience barriers that stem solely from the similarity of their first language and the language(s) of the host country.

2. METHOD

We used the restricted-use master microdata files of the PIAAC Canada assessment, with the nationally representative sample of 27,285 Canadians between 16 and 65 y.o.¹ Canada has two

¹The rules of using confidential Statistics Canada micro-data disallow reporting count data with 5 or less observations in a cell (to give a hypothetical example, the number of L1 speakers of Punjabi who took the test in French). Aggregated data are allowed for reporting as descriptive statistics and in regression models.

TABLE 1 | Sample sizes of immigrant groups by L1 and linguistic distance between L1 and English and French.

First language	Unweighted N	Distance from English	Distance from French
Chinese ^a	522	21	21
Dutch	24	6.40	14.55
English	711	0.00	13.27
French	126	13.27	0.00
German	66	6.87	15.02
Greek	15	10.77	14.50
Punjabi ^b	209	17.13	20.86
Italian	78	12.61	6.42
Polish	57	12.21	15.94
Portuguese	66	13.72	5.44
Spanish	226	13.77	5.49
Ukrainian	18	11.17	14.91

^aDistance values for Chinese are imputed.

^bDistance values for Punjabi are adopted from Gujarati.

official languages La (English and French) and thus the PIAAC instruments were administered in both languages. All scales of the PIAAC are psychometrically comparable between the English and French versions (and also across versions of the PIAAC assessments in all languages). Test reliability of both cognitive assessments (literacy, numeracy) and non-cognitive ones (e.g., reading habits, cultural capital) is extremely high. For numeracy, for instance, agreement between raters in the PIAAC numeracy task was 99.1% within-country and 96.7% across countries, see full details on reliability and validity in Organisation for Economic Co-operation and Development (2013).

The data-processing steps were as follows. We selected individuals who were born outside of Canada. Furthermore, we only considered those immigrants who reported one of the twelve L1s that received a specific language code in the PIAAC database as the languages most commonly spoken in Canada at the time of data collection (see **Table 1**): we do not consider languages coded as Other. We further removed from consideration responders with invalid or missing responses for one of the dependent or independent variables, detailed below. Since our analyses incorporated country-level variables (e.g., GDP), we filtered out immigrants from countries represented by fewer than 10 participants. The final resulting data set consisted of 2,018 responders, of which 1,930 completed the PIAAC assessments in English and 188 in French. A total of 29 different countries were represented: 29 among the English and 7 in the French test-takers.

2.1. Dependent Variables

This study quantifies proficiency in English or French as La through the PIAAC literacy score. PIAAC assessments of literacy make use of tasks that test comprehension, evaluation, and integration of words, sentences, and texts in authentic information-processing contexts resembling those arising at home and in the workplace (blogs, instructions, recipes etc.; PIAAC Literacy Expert Group, 2009; Trawick, 2017) (for a

demo see <https://www.oecd.org/skills/piaac/onlineassessmentdemo/>). Literacy scores range on a scale from 0 to 500, with the middle of the scale of 250 points and a standard deviation of 100 points. The scores are divided into six proficiency levels described in <http://www.piaac.ca/docs/PIAACTableaux/Literacy%20table.pdf>. We also considered scores in the numeracy assessment, defined over the same range. As with literacy, PIAAC construes numeracy broader than the ability to operate with numbers and rather tests the skills required for real contexts like “understanding purchases and receipts, reading maps, cooking, or engaging in home repairs” (Programme for International Assessment of Adult Competencies, 2012).

As the literature demonstrates (e.g., Adserà and Pytliková, 2016; Battershill and Kuperman, 2022), the potential impact of linguistic distance can propagate to the social sphere, constraining the immigrant’s willingness or ability to engage in civic, political, or cultural activities. We considered two of the measures of civic engagement reported in PIAAC: frequency of volunteering and political efficacy. Volunteering is assessed an ordinal response to the question “In the last 12 months, how often did you do voluntary work,” with five possible valid responses “Never (1), less than once a month (2), less than once a week but at least once a month (3), at least once a week but not every day (4), or every day (5).” Political efficacy is assessed as a Likert-scale response to the question “People like me don’t have any say about what the government does.” with the following valid response options “Strongly agree (1), agree (2), neither agree nor disagree (3), disagree (4), and strongly disagree (5).”

Economic standing of currently working responders was measured using their income, reported as Yearly income percentile rank with the following valid levels: <10 (1), 10 to <25 (2), 25 to <50 (3), 50 to <75 (4), 75 to <90 (5), and 90 or more (6).

A final dependent variable is self-reported health status defined as a response to the question “In general, would you say your health is excellent, very good, good, fair, or poor?” with options “Excellent (1), very good (2), good (3), fair (4), or poor (5).”

2.2. Independent Variables

Lexical distance is the variable of critical interest for this study. An excellent review of proposals on how to operationalize distance between language pairs is available (Schepens et al., 2013a). In line with Schepens et al. (2013b, 2020) and Isphording and Otten (2014), among others, this study makes use of lexical distance, i.e., a metric of dissimilarity between basic vocabularies of languages. Research on evolutionary change within a language family robustly demonstrates that languages that share a larger number of cognates, i.e., words that have a common historic origin, are more likely to have split from one another more recently (Swadesh, 1952) and thus can be represented closer to one another as branches on the family tree (see review for the Indo-European languages in Dyen et al., 1992). We used the estimates of lexical distance between Indo-European languages by Bouckaert et al. (2012), which are based on 5,995 cognate sets from Dyen et al. (1992). The estimates for all language pairs were derived from a phylogenetic language family tree and represent

the summed lengths of branches that connect both languages to each other (Schepens et al., 2013a). Thus, we obtained lexical distance estimates from all Indo-European languages in our set to English and to French (Table 1). Since Bouckaert et al. (2012) data did not list Punjabi, which is one of the common immigration languages specified in PIAAC, we substituted its lexical distance estimates by the estimates for a closely related Indo-Iranian language Gujarati. For the only non-Indo European language in our sample (Chinese), we followed the proposal by Schepens et al. (2020) and used a maximum (rounded) lexical distance observed in the Indo-European language family tree (see Table 1). For immigrants speaking English or French as their L1 and choosing the respective test language, lexical distance was zero; greater values of lexical distance indicated a greater dissimilarity between the immigrant’s L1 and the La of the test.

Other quantifications of linguistic distance rely on either morphological or phonological similarity in a language pair (Schepens et al., 2020). The main advantage of these metrics of linguistic distance over lexical distance is that they can apply across language families. Since 11 out of 12 languages in our sample belonged to the same Indo-European family, we confined ourselves to the more commonly used *lexical* representation of linguistic distance (e.g., Gray and Atkinson, 2003; Bouckaert et al., 2012). Also, arguably, the lexical distance is most appropriate for the literacy assessment that entirely relies on printed texts, tables, figures, and other visual means of representing information. One corollary of this bias toward visual language processing is that the present study does not shed light on some of the linguistic faculties critical for day-to-day communication ability, i.e., speaking or listening skill of immigrants or their rate of acquisition of spoken La.

Other independent variables were selected on the basis of the cross-national analyses of major predictors of literacy (Kyröläinen and Kuperman, 2021). Some variables are often considered as predictors of other dependent variables than literacy as well. They included gender (male 1, female 2), age (numeric), years since immigration (numeric), and the highest level of education completed by the individual and, separately, their mother’s education level [lower secondary or less (1), upper secondary (2), post-secondary—non-tertiary (3), tertiary—professional degree (4), tertiary—bachelor degree (5), or tertiary—master/research degree (6)]. We also included a measure of an individual’s cultural capital defined as the number of books that “were there in your home when you were 16 years old” [10 books or less (1), 11–25 books (2), 26–100 books (3), 101–200 books (4), 201–500 books (5), or more than 500 books (6)]. We also took into account the reported use of the reading skill at home (discretized into five quantiles). One of our questions was whether literacy fully mediates the effect of lexical distance on economic, social and health outcomes. Thus, literacy was used both as a dependent variable and a predictor in the models fitted to other dependent variables. The binary variable reflecting test language (English or French) did not reach significance in any analyses discussed below (and shown in the **Supplementary Materials**) and is not reported further.

In addition to the individual-level variables recorded in PIAAC, we considered characteristics of the immigrant's country of origin (Beenstock et al., 2001; Isphording, 2014; Schepens et al., 2020). These were the country's gross domestic product (GDP), schooling, i.e., the percent of population enrolled in secondary education, and life expectancy: All estimates came from the World Bank database. Since neither the country's schooling rate nor life expectancy reached significance in any of the models discussed in the remainder of the paper (and shown in the **Supplementary Materials**), we do not report these variables further.

2.3. Statistical Considerations

PIAAC data come with weights that enable each observed respondent to represent a larger segment of the Canadian population. As described in Kyröläinen and Kuperman (2021), we used ordinary least squares regressions with Jackknife Repeated Replication weights that correct for the complex design of the PIAAC samples (Organisation for Economic Co-operation and Development, 2013). The appropriate regression functions are implemented in the package `intsvy` that is designed specifically for the PIAAC data (Caro and Biecek, 2017) and is provided in the statistical platform R 3.6.1 (R Core Team, 2021). Specifically, regression models in this package implement a procedure where a regression model uses weights to estimate the effects of individual predictors, and the amount of variance explained by the regression model, for the entire population. Information-processing skills like literacy and numeracy are represented for each individual *via* a set of plausible values, rather than a single value: This setting reflects the fact that each participant only completes some of the assessment items. Thus, weighted regression models fitted to literacy and numeracy as the dependent variable accounted for variability of their plausible values, as implemented in function `intsvy.reg.pv`. To account for the effect of literacy as an independent variable, we calculated the mean of plausible values of literacy per participant and used function `intsvy.reg`. Regression models below report *t*-values: Effects associated with $|t| > 2.00$ were considered significant at the 5% threshold. Since no repeated measures were reported for the same person, we did not use mixed-effects models that can account for the by-participant and by-item variability.

The critical effect of lexical distance was evaluated in two ways, for its statistical significance in a regression model with multiple controls, and the amount of unique variance it explained over and above the controls. The latter quantity was measured as ΔR^2 , the difference in the explained variance between the full model that included lexical distance and the model without this predictor. Because weighted regressions in the `intsvy` package do not operate on ordinal data, we treated ordinal dependent variables as interval: These inferential estimates need to be approached with caution. Estimates of the national GDP were log-transformed to attenuate the influence of outliers.

3. RESULTS

This section reports analyses of information-processing skills, civic engagement, income, and health of 2,018 immigrants to

Canada representing 12 first languages, including English- and French-speaking immigrants. **Table 1** summarizes representation of languages and lexical distance from the two test languages, English and French.

3.1. Literacy and Numeracy

Lexical distance had a significant negative impact on individual literacy scores (**Supplementary Table 1**) [$\hat{\beta} = -0.544$; $SE = 0.256$; $t = -2.126$]. The unique contribution of lexical distance over and above known major predictors of literacy (e.g., education level, age, reading at home, national GDP) was fairly small [$\Delta R^2 = 0.4\%$]. Still, the estimated advantage in literacy scores between immigrants speaking English or French and immigrants speaking the language with the maximum distance from either test language (Chinese) is 11.4 points, or roughly one-quarter of the PIAAC literacy scale's standard deviation. This exceeds the estimated difference of 10.4 points in literacy scores between holders of a bachelor's degree (Education level 5) and graduate degree (Education level 6). If we gloss over the foreign-born speakers of English or French as their first language, the Chinese-speaking sample can be compared against Dutch (minimum lexical distance from English, 6.40) or Portuguese (minimum lexical distance from French, 5.44). In such comparisons, the estimated difference is around eight points, or roughly one-fifth of the standard deviation of the PIAAC literacy scale. This effect is comparable, for instance, to the difference between high-school diploma and tertiary degree as the highest educational attainment of the mother. Lexical distance did not produce a significant effect on numeracy scores [$\hat{\beta} = 0.276$; $SE = 0.270$; $t = 1.023$; $\Delta R^2 = 0.1\%$; see **Supplementary Table 2**].

3.2. Civic Engagement: Volunteering and Political Efficacy

There was a significant negative effect of lexical distance on volunteering frequency [$\hat{\beta} = -0.015$; $SE = 0.006$; $t = -2.716$; see **Supplementary Table 3**]. Furthermore, lexical distance made a substantial unique contribution of 1.0% to the amount of variance explained by the entire model, $R^2 = 10.0\%$. The effect was observed while adjusting for other major predictors of volunteering frequency. Notably, these predictors included literacy which showed a strong positive effect: Individuals with higher literacy scores reported a more frequent engagement on volunteering activities [$\hat{\beta} = 0.003$; $SE = 0.001$; $t = 3.388$; see Battershill and Kuperman, 2022]. Thus, lexical distance has a direct effect on this facet of civic engagement, besides indirectly affecting volunteering as a co-determiner of literacy.

Conversely, subjective political efficacy, or the evaluation whether one's voice can affect change in the country's governance, was not correlated with lexical distance when controlling for other predictors [$\hat{\beta} = -0.002$; $SE = 0.005$; $t = -0.314$; $\Delta R^2 = 0.0$; see **Supplementary Table 4**]. Literacy was positively correlated with the measure of political efficacy [$\hat{\beta} = 0.003$; $SE = 0.001$; $t = 2.774$, in line with Grotlüschen et al. (2016) and Battershill and Kuperman (2022), individuals with higher literacy scores considered their voice to be more influential in the country's governance.

3.3. Income

A total of 1,380 of individuals in our sample reported a valid, non-missing value for income. The weighted regression model fitted to the income measure did not reveal a unique effect of lexical distance [$\hat{\beta} = -0.006$; $SE = 0.006$; $t = -1.012$; $\Delta R^2 = 0.0$; see **Supplementary Table 5**]. Literacy had a strong positive effect on income [$\hat{\beta} = 0.004$; $SE = 0.001$; $t = 3.425$].

3.4. Health

Higher values of the subjective estimate of health, as coded in PIAAC, reflect poorer health (1 = excellent, 5 = poor). Lexical distance had a strong effect on the health measure: Speakers of languages less similar to the test language reported poorer subjective health, in line with Clarke and Isphording (2017). The effect size was substantial, as lexical distance uniquely explained 2.6% out of the total 11.1% of variance explained by the entire model (see **Supplementary Table 6**).

4. DISCUSSION

The linguistic literature has long since established that the degree of similarity between the first and the additional language of a speaker (L1 and La, respectively) determines the ease with which the speaker learns the La. This constraint on learnability—often coupled with practical limitations of learning time, exposure to La, incentives to learn La and other factors—may lead to a lower average proficiency in La among L1 speakers of languages that are more linguistically distant from La. The notion of linguistic distance has been embraced by the literature in labor economics, which positions proficiency in languages among critical components of human capital. Studies of immigration have further demonstrated the relevance of the L1-La linguistic distance for educational and professional attainment, employability, income, health, and social acculturation for adult and children immigrants.

This paper set out to contribute to the body of knowledge regarding the impact of linguistic distance on various spheres of the immigrant's life. This study was different from some previous research in a few ways. First, we considered immigrants to Canada, who may be generally skewed toward higher educational levels, stronger professional training and work experience, and language proficiency than immigrants to several other developed countries, due to the Canadian admission criteria, see the Section Introduction. We argued that—due to this skewness—the effects of linguistic distance on both La proficiency and other outcomes may be restricted in range. Second, while prior work mostly concentrated on a small selection of outcomes, we simultaneously considered the effect of linguistic distance on cognitive skills (literacy and numeracy), economic outcomes (income), civic engagement measures (volunteering frequency and perceived political efficacy) and health outcomes. Also, the present focus was on the objective measurement of literacy, i.e., proficiency with understanding, evaluating and using written texts in contexts approximating the information and communication demands of the home and workplace. This focus contrasts with prior

influential studies that correlated linguistic distance with either speaking proficiency or self-reported subjective proficiency in La.

This broad coverage of outcomes and relevant predictors was made available by the PIAAC Canada assessments of information-processing skills and rich socio-demographic questionnaires. The third point in which our study stands out is that PIAAC assessments were conducted in both official languages of Canada, English, and French, and thus we were able to consider the linguistic distance between numerous L1s and two different LAs. We adopted the lexical measurement of linguistic distance (e.g., Van der Slik, 2010; Schepens et al., 2013a, 2016; Isphording, 2014; Isphording and Otten, 2014) between the 12 most common languages spoken in Canada, coded in PIAAC data, and English and French as LAs. Finally, in all analyses, we examined whether lexical distance reached statistical significance and explained unique variance over and above a comprehensive range of individual- and country-level predictors identified as important in prior research. Where applicable, the set of predictors included literacy: This decision enabled us to determine whether lexical distance had a direct effect on, say, economic or health outcomes, or whether this effect was only exerted indirectly, through mediation of literacy. The ultimate goal was to refine the current understanding of the group-level barriers that cross-linguistic differences pose for the adaptation and acculturation of immigrants in the host country.

Our findings confirmed the unique and multi-faceted role of linguistic distance in shaping the lived experience of immigrants. Many of the findings confirm earlier reports often obtained from other national datasets or skills (but see e.g., Chiswick and Miller, 2005; Adsera and Ferrer, 2015), while others refine the existing body of knowledge. Perhaps, the most intriguing outcome of our analyses was that the lexical measure of linguistic distance exerted influence on some cognitive or social measures (e.g., literacy or volunteering frequency) but not on the closely related ones (numeracy or political efficacy). Below, we discuss all findings, including the discrepant ones, by groups of outcomes.

Literacy is the cognitive and information-processing skill in our set that is expected to reveal the strongest association with linguistic distance (Chiswick and Miller, 1995, 2005; Isphording and Otten, 2014; Schepens et al., 2016, 2020). Indeed, our analysis confirmed a weak but significant tendency for speakers of languages similar to the PIAAC test language to demonstrate higher literacy scores than those of less similar languages, i.e., an advantage of one-quarter of standard deviation between the extreme distances. This advantage is much less pronounced than prior reports suggest. For instance, Isphording (2014) analyzed literacy scores of immigrants to nine countries as a function of lexical distance between their first language and the language(s) of the host country, using data from the International Adult Literacy Survey (one of predecessors of the PIAAC assessment Clair, 2012). Isphording reports that the advantage of native-speaking immigrants over those speaking a lexically distant language exceed one standard deviation, on average, after adjusting for multiple factors that are considered

in this study as well. We attribute the discrepant effect sizes (one-quarter versus greater than one standard deviation) in otherwise similar studies to cross-national differences in the distribution of human capital among immigrants, dictated by the differences in linguistic, educational, occupational, and professional eligibility requirements for some or all immigration types. As we argued above, the preponderance of economic immigrants to Canada and the requirements they must satisfy prior to immigration may skew host language proficiency in this population to a more advanced level than observed in many other developed Western countries, as demonstrated in the 1- and especially 1.5-generation immigrants by Levels et al. (2017). We believe this somewhat restricted range to account for weaker effects of lexical distance. This finding points to the importance of supplementing international analyses of literacy and similar skills by the more nuanced and contextualized country-specific qualification of the immigrant composition that regards the country's immigration and integration policies (e.g., Ferrer et al., 2006; Levels et al., 2017).

Lexical distance did not affect numeracy in our data, contrary to literacy. Thus, no evidence was found to support that systematic differences in La learnability driven by L1 background influenced the individual's ability to operate and interpret mathematical information in a range of common tasks. This discrepancy among cognitive skills supports the long-standing notion that numeracy-related skills transfer more easily and fully upon a change to a linguistically different environment (Xu et al., 2017).

Another sphere of immigrant acculturation examined here concerned civic engagement, see the Introduction. Prior work reviewed in Adserà and Pytliková (2016) suggested that the L1-La linguistic distance may not only constrain one's learnability or proficiency attainment in La but also the willingness and ability to participate in the life of one's community, institutions, and society. Our findings revealed that speakers of more distant languages from English or French as La are less engaged in volunteering activities, i.e., the activities that are considered a gateway toward individual integration at the level of local community, building networks, and improving quality of life (Gottlieb and Gillespie, 2008; Barrett and Brunton-Smith, 2014; Battershill and Kuperman, 2022).

We did not find a direct effect of lexical distance on individual income, contra findings by Adserà and Ferrer (2015) based on the Canadian census (1991–2006) data on immigrant men. It is possible that this effect was fully mediated by literacy, which did affect income strongly, or other control predictors in the regression model. It may also be that the selection criteria for Canadian economic immigrants leveled off the cross-linguistic differences.

Finally, we confirmed prior reports of the linguistic distance effect on health outcomes (Clarke and Ispording, 2017). In our data, poorer health was more commonly reported by speakers of languages more distant from English or French. We interpret

this as a cumulative systematic effect of La learnability and proficiency on lower health literacy, lower willingness, and ability to access health services, and lesser engagement in the local community and society at large, partly due to more limited communication skills.

This study has its limitations. The sample of Canadian immigrants considered here is relatively small ($N = 2,018$) and only represents a fraction of languages spoken in Canada. Similarly, some of the variables known to affect La acquisition—namely, the age at which acquisition has started and the nature of La use at home and workplace—are only partially available in the PIAAC data. We are expecting the upcoming second cycle of the PIAAC data collection to expand the sample and add both to the statistical power of the regression models and the range of linguistic distances and dependent variables. The questionnaires of PIAAC will also be enhanced to supply some of the data currently lacking.

In sum, the present data demonstrate the breadth of repercussions that linguistic distance between the immigrant's L1 and the host country's La have across multiple domains of life. Going beyond proficiency in La as determined by skill assessments, this study confirms that linguistic distance co-determines economic standing and prospects, acculturation and assimilation in the host society. The novelty of this contribution is in quantifying the impact of linguistic distance (or lack thereof) over a greater number of domains than covered in much prior work, using a rich nationally representative dataset of PIAAC Canada. Our analyses refine the current understanding of the interplay between lexical distance and literacy in co-determining cognitive, socio-economic, and health outcomes among linguistically diverse immigrants to Canada.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The paper uses microdata files of the Statistics Canada component of the Programme for International Assessment of Adult Competencies. The files are protected and cannot be distributed. I received a special permission to use the microdata. Requests to access these datasets should be directed to Research Data Center at McMaster University, rdc@mcmaster.ca.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This work was supported by the Social Sciences and Humanities Research Council of Canada Partnered Research Training Grant, 895-2016-1008, (Dr. Gary Libben, PI), the Canada Research Chair (Tier 2; Kuperman, PI), and the CFI Leaders Opportunity

Fund (Kuperman, PI). Thank are due to Kaitlyn Battershill for proof-reading and editing this manuscript and to the staff of the McMaster Research Data Centre for technical support. Thanks are also due to Job Schepens for providing estimates of linguistic distance.

REFERENCES

- Adsera, A., and Ferrer, A. M. (2015). "The effect of linguistic proximity on the occupational assimilation of immigrant men in Canada," in *IZA Discussion Paper*, No. 9499 (Bonn: Institute for the Study of Labor). doi: 10.2139/ssrn.2690747
- Adserá, A., and Pytlíková, M. (2016). "Language and migration," in *The Palgrave Handbook of Economics and Language*, eds V. Ginsburgh and S. Weber (Houndmills; Basingstoke: Palgrave Macmillan), 342–372. doi: 10.1007/978-1-137-32505-1_13
- Barrett, M., and Brunton-Smith, I. (2014). Political and civic engagement and participation: towards an integrative perspective. *J. Civil Soc.* 10, 5–28. doi: 10.1080/17448689.2013.871911
- Battershill, K., and Kuperman, V. (2022). A bird's eye view of civic engagement and its facets: canonical correlation analysis across 34 countries. *PLoS ONE*.
- Beenstock, M., Chiswick, B. R., and Repetto, G. L. (2001). The effect of linguistic distance and country of origin on immigrant language skills: application to Israel. *Int. Migrat.* 39, 33–60. doi: 10.1111/1468-2435.00155
- Belot, M., and Ederveen, S. (2012). Cultural barriers in migration between OECD countries. *J. Popul. Econ.* 25, 1077–1105. doi: 10.1007/s00148-011-0356-x
- Bleakley, H., and Chin, A. (2008). What holds back the second generation? The intergenerational transmission of language human capital among immigrants. *J. Hum. Resour.* 43, 267–298. doi: 10.1353/jhr.2008.0028
- Borjas, G. J., and Chiswick, B. R. (2019). *Foundations of Migration Economics*. Oxford: Oxford University Press. doi: 10.1093/oso/9780198788072.001.0001
- Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S. J., Alekseyenko, A. V., Drummond, A. J., et al. (2012). Mapping the origins and expansion of the Indo-European language family. *Science* 337, 957–960. doi: 10.1126/science.1219669
- Bynner, J. (2004). Literacy, numeracy and employability: Evidence from the British birth cohort studies. *Liter. Numer. Stud.* 13, 31–48. doi: 10.4324/978020388889
- Caro, D. H., and Biecek, P. (2017). *intsvy: an R package for analyzing international large-scale assessment data*. *J. Stat. Softw.* 81, 1–44. doi: 10.18637/jss.v081.i07
- Casey, T., and Dustmann, C. (2008). Intergenerational transmission of language capital and economic outcomes. *J. Hum. Resour.* 43, 660–687. doi: 10.1353/jhr.2008.0002
- Chiswick, B. R., and Miller, P. W. (1995). The endogeneity between language and earnings: international analyses. *J. Labor Econ.* 13, 246–288. doi: 10.1086/298374
- Chiswick, B. R., and Miller, P. W. (1998). English language fluency among immigrants in the United States. *Res. Labor Econ.* 17, 151–200.
- Chiswick, B. R., and Miller, P. W. (2003). The complementarity of language and other human capital: immigrant earnings in Canada. *Econ. Educ. Rev.* 22, 469–480. doi: 10.1016/S0272-7757(03)00037-2
- Chiswick, B. R., and Miller, P. W. (2005). Linguistic distance: a quantitative measure of the distance between English and other languages. *J. Multiling. Multicult. Dev.* 26, 1–11. doi: 10.1080/14790710508668395
- Chiswick, B. R., and Miller, P. W. (2012). Negative and positive assimilation, skill transferability, and linguistic distance. *J. Hum. Capit.* 6, 35–55. doi: 10.1086/664794
- Clair, R. S. (2012). The limits of levels: understanding the international adult literacy surveys (IALS). *Int. Rev. Educ.* 58, 759–776. doi: 10.1007/s11159-013-9330-z
- Clarke, A., and Ispording, I. E. (2017). Language barriers and immigrant health. *Health Econ.* 26, 765–778. doi: 10.1002/hec.3358
- Crossman, E., Hou, F., and Picot, G. (2021). *Are the Gaps in Labour Market Outcomes Between Immigrants and Their Canadian-Born Counterparts Starting to Close?* Ottawa, Canada: Statistics Canada. 1.
- Dyen, I., Kruskal, J. B., and Black, P. (1992). An indoeuropean classification: a lexicostatistical experiment. *Trans. Am. Philos. Soc.* 82, 3–132. doi: 10.2307/1006517
- Ferrer, A., Green, D. A., and Riddell, W. C. (2006). The effect of literacy on immigrant earnings. *J. Hum. Resour.* 41, 380–410. doi: 10.3368/jhr.XLI.2.380
- Finnie, R., and Meng, R. (2007). *Literacy and Employability*, Statistics Canada. Available online at: <http://www.statcan.gc.ca/pub/89-604-x/89-604-x2011001-eng.pdf>
- Ginsburgh, V., and Weber, S. (2020). The economics of language. *J. Econ. Literat.* 58, 348–404. doi: 10.1257/jel.20191316
- Gottlieb, B. H., and Gillespie, A. A. (2008). Volunteerism, health, and civic engagement among older adults. *Can. J. Aging* 27, 399–406. doi: 10.3138/cja.27.4.399
- Gray, R. D. and Atkinson, Q. D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426, 435–439. doi: 10.1038/nature02029
- Grotlüschen, A., Mallovs, D., Reder, S., and Sabatini, J. (2016). "Adults with low proficiency in literacy or numeracy," in *OECD Education Working Papers*, No. 131 (Paris: OECD Publishing).
- Immigration Refugees Citizenship Services (2021). *Six Selection Factors—Federal Skilled Worker Program (Express Entry)*. Available online at: <https://www.canada.ca/en/immigration-refugees-citizenship/services/immigrate-canada/express-entry/eligibility/federal-skilled-workers/six-selection-factors-federal-skilled-workers.html> (accessed December 17, 2021).
- Ispording, I. E. (2014). Disadvantages of linguistic origin—evidence from immigrant literacy scores. *Econ. Lett.* 123, 236–239. doi: 10.1016/j.econlet.2014.02.013
- Ispording, I. E., and Otten, S. (2014). Linguistic barriers in the destination language acquisition of immigrants. *J. Econ. Behav. Organ.* 105, 30–50. doi: 10.1016/j.jebo.2014.03.027
- Jarvis, S., and Pavlenko, A. (2008). *Crosslinguistic Influence in Language and Cognition*. New York, NY: Routledge. doi: 10.4324/9780203935927
- Kellerman, E. (1979). Transfer and non-transfer: where we are now. *Stud. Second Lang. Acquis.* 2, 37–57. doi: 10.1017/S0272263100000942
- Kyröläinen, A.-J., and Kuperman, V. (2021). Predictors of literacy in adulthood: evidence from 33 countries. *PLoS ONE* 16, e0243763. doi: 10.1371/journal.pone.0243763
- Levels, M., Dronkers, J., and Jencks, C. (2017). Contextual explanations for numeracy and literacy skill disparities between native and foreign-born adults in western countries. *PLoS ONE* 12, e0172087. doi: 10.1371/journal.pone.0172087
- Marschak, J. (1965). Economics of language. *Behav. Sci.* 10, 135–140. doi: 10.1002/bs.3830100203
- Mayda, A. M. (2010). International migration: a panel data analysis of the determinants of bilateral flows. *J. Popul. Econ.* 23, 1249–1274. doi: 10.1007/s00148-009-0251-x
- Muñoz, C., and Muñoz, C. (2006). *Age and the Rate of Foreign Language Learning*, Vol. 19. Clevedon: Multilingual Matters Clevedon.
- Organisation for Economic Co-operation and Development (2013). *Technical Report of the Survey of Adult Skills (PIAAC)*. Organisation for Economic Co-operation and Development.
- PIAAC Literacy Expert Group. (2009). "PIAAC literacy: A conceptual framework," in *OECD Education Working Papers*, No. 34 (Paris: OECD Publishing). doi: 10.1787/220348414075
- Programme for International Assessment of Adult Competencies (2012). *Piaac in Canada*. Available online at: <http://www.piaac.ca/> (accessed December 17, 2021).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpos.2022.874195/full#supplementary-material>

- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Ringbom, H. (2006). *Cross-Linguistic Similarity in Foreign Language Learning*. Clevedon: Multilingual Matters. doi: 10.21832/9781853599361
- Schepens, J., Van der Slik, F., and Van Hout, R. (2013a). "The effect of linguistic distance across Indo-European mother tongues on learning Dutch as a second language," in *Approaches to Measuring Linguistic Differences*, eds L. Borin and A. Saxena (Berlin: De Gruyter Mouton), 199–230. doi: 10.1515/9783110305258.199
- Schepens, J., van der Slik, F., and van Hout, R. (2013b). Learning complex features: a morphological account of L2 learnability. *Lang. Dyn. Change* 3, 218–244. doi: 10.1163/22105832-13030203
- Schepens, J., Van der Slik, F., and Van Hout, R. (2016). L1 and L2 distance effects in learning L3 Dutch. *Lang. Learn.* 66, 224–256. doi: 10.1111/lang.12150
- Schepens, J., van Hout, R., and Jaeger, T. F. (2020). Big data suggest strong constraints of linguistic similarity on adult language learning. *Cognition* 194, 104056. doi: 10.1016/j.cognition.2019.104056
- Statistics Canada. (2016). *Census of Population. Technical Report, Statistics Canada Catalogue No. 98-400-X2016202*. Ottawa, ON: Statistics Canada. doi: 10.1787/ins_stats-2015-9-en
- Swadesh, M. (1952). Lexico-statistic dating of prehistoric ethnic contacts: with special reference to North American Indians and Eskimos. *Proc. Am. Philos. Soc.* 96, 452–463.
- Trawick, A. R. (2017). *Using the PIAAC Literacy Framework to Guide Instruction: An Introduction for Adult Educators*. Washington, DC: PIAAC.
- Van der Slik, F., Hout, R. v., and Schepens, J. (2019). The role of morphological complexity in predicting the learnability of an additional language: the case of la (additional language) Dutch. *Second Lang. Res.* 35, 47–70. doi: 10.1177/0267658317691322
- Van der Slik, F. W. (2010). Acquisition of Dutch as a second language: the explanative power of cognate and genetic linguistic distance measures for 11 west European first languages. *Stud. Second Lang. Acquis.* 32, 401–432. doi: 10.1017/S0272263110000021
- Van der Slik, F. W., Van Hout, R. W., and Schepens, J. (2015). The gender gap in second language acquisition: Gender differences in the acquisition of Dutch among immigrants from 88 countries with 49 mother tongues. *PLoS ONE* 10, e0142056. doi: 10.1371/journal.pone.0142056
- Xu, L., Zhong, J., and Maslov, A. (2017). *Skills Proficiency of Immigrants in Canada: Findings from the Programme for the International Assessment of Adult Competencies (PIAAC)*. Council of Ministers of Education.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Kuperman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.