Author/s:
Clarke, A;Isphording, IE

Title:
Language Barriers and Immigrant Health

Date:
2017-06-01

Citation:
Clarke, A. & Isphording, I. E. (2017). Language Barriers and Immigrant Health. Health Economics United Kingdom, 26 (6), pp.765-778. https://doi.org/10.1002/hec.3358.

Persistent Link:
https://hdl.handle.net/11343/291341

Language Barriers and Immigrant Health

Andrew Clarke[*] and Ingo E. Isphording [†]

February 10, 2016

Abstract

We study the impact of language deficiency on the health status of childhood migrants to Australia. Our identification strategy relies on a quasi-experiment comparing immigrants arriving at different ages and from different linguistic origins. In the presence of considerable non-classical measurement error in self-reported language proficiency, our results provide lower and upper bounds for a strong negative effect of English deficiency on health of between one half and a full standard deviation in the health score.

Keywords: International Migration; Language Skills; Health.
JEL Classification: F22, I12, J24, J61

[*] Department of Economics, University of Melbourne, Level 4, Faculty of Business and Economics Building, 111 Barry Street, Carlton, Victoria, Australia; andrew.clarke@unimelb.edu.au.

[†] Corresponding author: Ingo E. Isphording, Institute for the Study of Labor, Schaumburg-Lippe-Str. 5-9, 53113 Bonn, Germany. Phone, +49 (228) 3894-204, Fax +49 (228) 3894-510, E-mail isphording@iza.org.

1 Introduction

Good health outcomes for the foreign born are both a requisite and an outcome of successful integration into their host country society and at an aggregate level shape the net fiscal impact of immigration. A crucial channel through which the health of immigrants is affected during their residence in the host country is the acquisition of the destination language. Despite a large epidemiological literature providing empirical evidence for the association of language barriers and low literacy with inferior health outcomes (see De Walt et al. (2004) and Eichler et al. (2009) for systematic overviews), there is only limited evidence on the causal mechanisms.

This study provides evidence for the causal effect of host country language skills on the health status of immigrants. Framed in terms of a Grossman-style health production model, causal pathways between language deficiency and health may be summarised in terms of two main effects. First, language proficiency affects the access to inputs into health production such as access to jobs that pay higher wages and involve greater safety. Second, language skills are also associated with improvements in the efficiency of health production by raising the marginal productivity of the inputs into health production.

In order to identify the causal effect of language proficiency on health in the presence of endogeneity and measurement error in language skills, we construct a quasi-experiment based on a systematic decay in the ability to learn new languages during childhood. Immigrants arriving after a certain age threshold (the 'critical period') have to invest significantly more effort than young arrivals to reach comparable levels of host country language proficiency. Moreover, this increase in the required investment of effort depends on the dissimilarity between the home and host country language. Relative to immigrants whose first language is linguistically distant to the destination language, linguistically close immigrants face a lower increase in the costs of acquisition by a later age at migration (Isphording, 2014).

These differences by age at arrival and linguistic distance provide the necessary variation to apply an identification strategy similar to Bleakley and Chin (2004). We employ an instrument that is based on a simple difference-in-difference exercise, comparing immigrants with different ages at arrival and with differing linguistic distance to the host country language. A distinctive feature of our approach is the use of a continuous measure of linguistic dissimilarities between home and host-country language which allows us to control for the considerable heterogeneity within the group of non-native speaking immigrants. This identification strategy allows us to construct an instrument that is arguably affecting immigrant health only through differences in the costs of language acquisition.

Our empirical analysis utilizes a large individual level dataset for Australia, the Household, Income and Labour Dynamics in Australia (HILDA) Survey, that provides a quasi-objective measure of health, self-reported language skills, and an extensive range of socioeconomic and migrant background characteristics. There are several compelling reasons that motivate the study of the relationship between language skills and health within Australia. First, Australia is a large immigrant receiving country, in 2013 almost 28% of the Australian population was born overseas. Australia also operates a binding and tightly controlled immigrant selection policy requiring applicants to satisfy minimum health requirements. Second, Australia has a public health system that provides subsidised payments for medical services and prescription medications. Consequently, differences in health status associated with differences in language proficiency are more likely to reflect differences in the utilisation of health services rather than access to health services. Third, Australia has a substantial stock of immigrants with an English mother tongue who act as a large control group for identifying the causal effects of language proficiency on health.

Our results provide evidence for a large detrimental causal effect of language deficiency on the (physical) health status of immigrants. While the OLS estimates indicate a health disadvantage of about half a standard deviation in physical health scores for deficient speakers, the IV estimates are approximately twice as large.[1] This main causal effect provides evidence for a substantial impact of language skills on the health stock of immigrants, particularly when compared to the much studied education-health gradient for native-born individuals. For example, the unconditional mean difference in health between native-born university educated individuals and native-born high school dropouts is only slightly less than one half of a standard deviation in the health score.

As measurement error in subjective information on language proficiency is likely affecting our OLS and IV results in different directions, inducing a downward attenuation bias in the OLS results and potentially inflating our IV results by non-classical measurement error, our results provide lower and upper bounds for the effect of language deficiency on the health scores. We use an auxiliary dataset containing both subjective and objective information on language skills to provide secondary evidence on the importance of this non-classical misclassification error in self-reported language proficiency.

2 Identification strategy

To establish the causal pathways between the language skills of immigrants and their health status, it is useful to apply a standard Grossman-style health production model (Grossman, 1972). In this model, individual health is a depreciating capital stock dependent on an initial endowment that can be restocked by a health production process taking into account health investments, expenditures and constraints.

The influence of language proficiency might play a similar role as the much ana-

[1] Guven and Islam (2015) use a related identification strategy for an extremely large number of different labour market and socioeconomic outcomes, including health as one of many outcomes. They provide quite large estimates (almost two standard deviations in the health score) for the effect of English proficiency on the health of female immigrants and a large yet imprecise estimate for male immigrants, suggesting lower health for male English proficient immigrants. The unusually large effects and surprising gender differences can be shown to be largely driven by failing to account for the selection of immigrants according to initial (childhood) health conditions and parental education which are both potentially related to their instrument. More recently, Aoki and Santiago (2015) analyze the causal link between language skills, education, fertility and health for immigrants in the UK, and find no significant causal link between language skills and self-reported health status.

lyzed effect of education on health (for overviews, see Grossman and Kaestner (1997) and Cutler and Lleras-Muney (2006)). First, language proficiency has an indirect role by affecting the vector of inputs into the health production function, such as employment, occupational choice, wages, and social integration. Second, language proficiency has a direct effect by improving the efficiency of health production. This efficiency gain might be achieved by increased productive efficiency, raising the marginal productivity of health inputs. Alternatively, the efficiency gain could accrue through an increased allocative efficiency, allowing a more efficient combination of the inputs to health production. These efficiency gains accrue through a greater ability to interact with the health care system, including direct communication with health care personnel, but also access to and the processing of health information provided in magazines, booklets or public information campaigns. A dearth in both the quantity and quality of health information might then affect the health stock of language deficient immigrants.

Both these direct and indirect links imply an unambiguously negative effect of language deficiency on health outcomes that could be naively modeled with a simple reduced form model where the adult health of immigrant i from source country c in time t (H_{ict}) is a linear function of English deficiency (EDEF_{ict}), age at arrival (AAI_{ic}), the linguistic distance between English and the linguistic background of an immigrant (LD_{ic}), and other economic and demographic characteristics ($\mathbf{X}_{ict}$):

$$\text{H}_{ict} = \beta_0 + \beta_1 \,\text{EDEF}_{ict} + \beta_2 \,\text{AAI}_{ic} + \beta_3 \,\text{LD}_{ic} + \mathbf{X}_{ict}^T \gamma + \varepsilon_{ict}. \tag{1}$$

Several sources of endogeneity raise doubts on a causal interpretation of β_1. First, it is likely that unobserved variables that are correlated with both health and language skills generate an omitted variable bias. Second, since lower health decreases expected wages and lowers the incentives to invest in host-country specific human capital, we might expect causality from health to language skills. Third, in the absence of objective test scores for language fluency, we are restricted to a self-reported measure of English deficiency. The proclivity of such a self-reported measure to measurement error is well-documented (Charette and Meng, 1994; Dustmann and van Soest, 2001; de Coulon and Wolff, 2007), leading to an under-estimation of the true effect of language skills in naive models estimated by OLS.[2]

To address these concerns, we employ a twofold identification strategy following Dustmann and van Soest (2002). First, we take advantage of the longitudinal dimension in our data and replace the binary English deficiency indicator EDEF_{ict} with predictions from a regression of the binary English deficiency indicator on a non-linear time trend (by including period dummies) to reduce the amount of time-variant noise, analogous to using moving averages to override noisy measurements.[3]

Second, we develop an instrumental variable based on age at immigration and linguistic background to correct for time-invariant measurement error following the approach by Bleakley and Chin (2004). This strategy uses the systematic and sudden decrease in the ability to acquire new languages around an age threshold commonly labeled as the 'critical period'. Reasons for this structural break have been documented as being neuro-biologically based (Newport, 2002), but can also be rooted in psychological and social factors that favor younger children, such as a greater motivation and pressure to use the second language on a daily basis (McLaughlin, 1992). As this critical period in language acquisition might coincide with further critical periods in youth development (van den Berg et al., 2014), we cannot simply use this structural break as identifying variation. Rather, identification is based on differences in the effect of the age at arrival between different linguistic groups ranked by their dissimilarity to English. This results in a difference-in-difference exercise, in which we compare immigrants from close and distant linguistic background arriving before and after this critical period.

Our main identifying assumption is that differences in linguistic origin raise the costs of language acquisition, but have no further direct or indirect effect on health

[2]Dustmann and van Soest (2001); Bleakley and Chin (2004); Chiswick and Miller (1995)

[3]This adjustment has a non-negligible effect on the OLS results later reported in Table 5. OLS regression results without the adjustment indicate an approximately 40% lower partial correlation between English deficiency and the health scores, consistent with attenuation bias. In contrast, the IV estimates with this adjustment are virtually identical to the IV estimates that do not use this adjustment.

production (conditional on $\mathbf{X}_{ict}$). Consequently, differences in health outcomes between linguistic groups can be solely attributed to these differences in the costs of language acquisition. Comparing across different linguistic groups and different ages at arrival enables us to partial out non-linguistic influences on age at arrival (such as earlier acculturation, earlier access to host country educational institutions, health care and diet) from the effect of age at arrival on language acquisition.

Figure 1 clarifies our identification strategy, showing a simple comparison of conditional means of health scores and language deficiency by age at arrival and native language . While the language acquisition of native-speaking immigrants is unaffected by their age at arrival (dark-grey line), the probability of being deficient as an adult is strongly increasing by age at arrival for immigrants with a non-English linguistic background (Panel A). This pattern of differences between native-speaking and non-native-speaking immigrants seems to be transmitted into an analogous pattern in their health status. While health scores are marginally increasing by age at arrival for native speakers, non-native speaking immigrants seem to suffer from significant health disadvantages when they arrive at later ages.

A distinctive feature of our identification approach is the use of a continuous measure of linguistic dissimilarities between home and host-country language which allows us to control for the considerable heterogeneity within the group of non-native speaking immigrants. Immigrants from European countries share a certain amount of language similarity with English-speaking immigrants. For example, Scandinavian and German languages originate in the Germanic language family, but also Italian and Greek languages come from an Indo-European language origin. Asian immigrants do not share these language family ties with English.

We estimate a first stage explaining English deficiency as a function of age at immigration, linguistic background LD_i (varying by country of birth and whether an immigrant learned English as first language) and the interaction between both variables:

$$\widetilde{\mathrm{EDEF}}_{ict} = \gamma_0 + \gamma_1\,\mathrm{AAI}_{ic} + \gamma_2\,\mathrm{LD}_{ic} + \gamma_3\,\{\mathrm{AAI}_{ic} \times \mathrm{LD}_{ic}\} + \mathbf{X}_{ict}^T\,\phi + \omega_{ict} \qquad (2)$$

Given our identifying assumption, that differences in linguistic origin affect health production only through increasing the costs of language acquisition, and controlling for a rich set of individual characteristics, the interaction term $\{\mathrm{AAI}_{ic} \times \mathrm{LD}_{ic}\}$ will be uncorrelated with the unobservable determinants of adult health. Consequently, estimating equation (2) will provide exogenous predictions $\widehat{EDEF}_{ict}$ that we can use to explain health outcomes in the second stage, excluding the instrument:

$$\mathrm{H}_{ict} = \beta_0 + \beta_1 \,\widehat{\mathrm{EDEF}}_{ict} + \beta_2 \,\mathrm{AAI}_{ic} + \beta_3 \,\mathrm{LD}_{ic} + \mathbf{X}_{ict}^T \gamma + \mu_{ict} \tag{3}$$

We estimate this equation using 2SLS, with standard errors clustered at the individual level to account for repeated observations of the same individual in the HILDA panel. We run a number of nested specifications, with our preferred specification including time and country-of-origin fixed effects and a large set of control variables measured at the individual and regional level, as outlined in the next section.

Importantly, the HILDA data allows us to alleviate concerns about confounding and unobservable differences in initial health status: Immigrants who arrive at a younger age have a shorter exposure to potentially adverse health conditions in their source countries. If these childhood health conditions differ systematically between immigrants from different linguistic backgrounds, initial health status could potentially be correlated with both our instrument and adult health status. By additionally controlling for a subjective measure of childhood health in all specifications we ensure the validity of our exclusion restriction.

By construction, measurement error in a binary variable is negatively correlated with the 'true' value of the binary variable. Individuals reporting to have English deficiency can only deviate towards English non-deficiency while individuals reporting not to have difficulties can only deviate towards English deficiency. In this case, the OLS estimates will be downward biased and take the same sign as the 'true' effect provided there are not too many 'false positives' (individuals with good language skills reporting language deficiencies). This form of non-classical measurement error will also lead to first stage estimates in the IV regressions that are biased towards zero and subsequently inflated reduced form IV estimates, as shown by Brachet (2008). Consequently, the OLS estimates provide a lower bound on the effect of language skills on health while the IV estimates provide an upper bound. Later, we use auxiliary information from the Australian Adult Literacy and Lifeskills survey (ALLS) which contains both subjective and objective measures of language skills to assess the extent of this non-classical measurement error.

3 Data

Our data consists of a sample of childhood migrants who arrived in Australia before the age of 18 and who are aged between 18 and 65 at the point of the interview, drawn from the Household, Income and Labour Dynamics in Australia (HILDA) Survey 2001-2011. Quasi-objective health scores are measured using the SF–36, a commonly applied health questionnaire based on 34 items for different domains of health. The single items are grouped into 4 different domains of physical health: physical functioning, role limitations due to physical problems, bodily pain, and general health perceptions. From these sub-domains a physical health score is derived using a standardized coding scheme. The physical health score is defined on a 0–100 range, with a sample mean of 78.5 points and a standard deviation of 19.9 points.

Our main variable of interest is the English language proficiency of the individual which is measured in our data as their current self-reported ability to speak English. We construct a binary indicator of English deficiency that takes on the value of 0 if the respondent reports speaking English 'very well' (the highest category) and a value of 1 otherwise.

We further utilise a large number of control variables at the individual level (gender, age in categories of five years, and parental education), regional level (state-fixed effects, urban environment, the SEIFA index of local socio-economic advantage and disadvantage), and a full set of country-of-origin indicators.[4] A set of cohort (decade of arrival) indicators are used to control for potentially unobservable changes the selection of immigrants.

As an additional health indicator, the HILDA data offers information on subjective childhood health status before the age of 15. As noted above, controlling for individual differences in childhood health might be crucial for identification—not controlling for childhood health might overstate the importance of English deficiency on adult health by attributing part of the detrimental effect of prolonged exposure to adverse environments during childhood to English deficiency.

We augment the HILDA data with a measure for linguistic dissimilarities between the mother tongue of an immigrant and English.[5] The Automatic Similarity Judgement Program (ASJP) of the German Max Planck Institute of Evolutionary Anthropology (Bakker et al., 2009) offers a convenient way to summarize the various dimensions of language dissimilarities (such as vocabulary, pronunciation, phonetic inventories, grammar etc.) within a single summary statistic.

The measurement relies on an automatic comparison of a list of words chosen for their universal availability, referred to as the Swadesh-list (upper panel of Table 1). Word pairs between languages are transcribed into phonetic script and the minimum number of sounds needed to transfer the one word into the other is calculated (the 'Levenshtein distance'). Examples for this calculation between English and German are provided in the lower panel of Table 1: While the German and English word

[4]In the presence of a considerable negative correlation between parental education and children's age at immigration (Clarke, 2015), it is crucial to control for parental education in order to support our identifying assumption. A full set of country of origin fixed effects are used to control for time-invariant confounding variables, such as differences in the pre-migration exposure to English.

[5]The model is still identified if linguistic distance is assigned solely on the basis of country of birth. However, this would impart some measurement error in our measure of linguistic distance for multi-lingual countries with English as an alternative language. In light of this, for individuals who do not report English as their mother tongue, we approximate their (non-English) mother tongue by the predominant language of their country of birth. Immigrants from non-English-speaking countries who report English as their first language are accordingly assigned a linguistic distance of zero. Nonetheless, assigning linguistic distance by predominant language of a country for individuals with a non-English mother tongue inevitably introduces measurement error into the variable. In the robustness checks in the appendix, we show that excluding multilingual countries does not considerably affect our findings.

for fish are similarly pronounced (resulting in a minimum distance of 0), the English *mountain* and the German word *Berg*, transcribed in phonetic script as *maunt3n* and *bErk*, have no similarities, resulting in a minimum distance of seven. Finally, a normalized average is taken over the different word pairs to derive the summary measure of linguistic distance.

The resulting continuous and cardinal index is intended to capture differences in the initial hurdles raised by linguistic barriers in the language acquisition of immigrants (Isphording and Otten, 2013). This measure of linguistic distance behaves well as a predictor for expert opinions on language relations and differences in vocabulary and grammar (Wichmann et al., 2010). In our HILDA sample, the closest languages to English spoken by immigrants are Dutch, Swedish and Westvlaams (Belgium). In contrast, Vietnamese, Somali and Khmer (Cambodia) represent the most distant languages.[6]

Our final sample consists of 5,706 person-year observations, including 4,420 observations with English as mother tongue (first language learned), and 1,286 observations with a foreign linguistic background, based on 569 native speakers and 278 non-native speakers. Table A.1 in the appendix provides an overview of the sample composition by country of origin. Among the English-speaking immigrants, the vast majority come from the United Kingdom and New Zealand, while the group of immigrants with a foreign linguistic background have diverse ancestries mainly from Asia and Europe.

Table 2 provides summary statistics by linguistic origin and age at arrival. Based on the unconditional means, no striking systematic differences in health scores can be observed, although scores seem to be marginally higher for late arrivers (after age of 12) for native-speaking immigrants, while they are marginally lower for non-native speakers arriving late. Unsurprisingly, the main difference between the samples can be observed in the probability of being deficient in the English language. While the probability of being deficient is negligible for native speakers, around 7% of the young

[6] Linguistic distance is potentially correlated with further unobservable differences between countries, such as 'cultural distance'. In the robustness checks presented in the appendix, we show that our results are robust towards controlling for several different proxy variables for cultural dissimilarities between countries.

arriving non-native-speaking immigrants report to have problems in communicating in English. This share is increasing to almost a third (35%) for late arrivers who did not learn English as their first language.

Comparing within-person changes in English deficiency over time, we observe approximately as many transitions from English deficiency to English non-deficiency as from non-deficiency to deficiency status. Given the fact that we observe most immigrants in the later stages of their assimilation (the average immigrant is observed 33 years after immigration, we reasonably assume that it is highly unlikely that acquired language skills are deteriorating as fast as the rate at which immigrants are acquiring new language skills. Analogous to a similar discussion in Dustmann and van Soest (2002), we interpret this as suggestive evidence for presence of considerable measurement error our English deficiency variable.

To further assess the extent of the non-classical measurement error in self-reported language skills, we utilize auxiliary data from the Australian version of the Adult Literacy and Life Skills Survey (ALLS). For the purposes of this paper, only the single skill domain prose literacy is used (the knowledge and skills needed to understand and use various kinds of information from text including editorials, news stories, brochures and instructions manuals). Proficiency is measured along a continuous scale ranging from 0 to 500.

4 Results

Ordinary least squares results

We start our analysis by first providing a descriptive overview of the relationship between adult health status, language deficiency and the age at arrival of childhood immigrants. In Table 3 we analyze the relationship between health and age at arrival. Restricting the age at arrival profile to be identical for native and non-native speakers, there is no evidence that age at arrival is related to inferior health outcomes (Panel A). In contrast, we find a very distinctive negative relationship between age at arrival and the health outcomes for non-native speakers when we allow the coefficients to differ between native and non-native speakers. The estimated pattern is insensitive to the inclusion of potential additional confounding factors included in the subsequent columns. For the latest arrival cohort (arriving at ages between 15 and 17), the average difference in health between a native and a linguistically distant speaker (with $LD = 1$) amounts for almost a standard deviation of the physical health score.

As the principal difference between native and non-native-speaking immigrants is likely to be found in their language skills, we now examine the relationship between age at arrival and the incidence of language deficiency in Table 4. Similar to the health scores, there is no systematic relationship between age at arrival and the likelihood of English deficiency for native speakers. Consistent with the critical period hypothesis, the likelihood of being deficient in English increases strongly with age at arrival for non-native speakers: Linguistically distant immigrants in the latest age at arrival cohort display a difference of 36.9 probability points of being deficient in English, compared to English-speaking immigrants.

Instrument variable results

Using the identification strategy outlined in Section 2, Table 5 presents two stage least squares estimates of the causal effect of language skills on health. The interaction term between linguistic distance and age at arrival is highly significant in the first stage (column (1)). OLS estimates indicate a moderate negative correlation between English deficiency and the health scores (column 2). English deficiency is associated with a lower health score by about 11.7 points, about half a standard deviation in the health score. The IV results in columns (3) - (6) strongly exceed these (naive) OLS results. Being deficient in English causes a lower health score by about 20 points, approximately one standard deviation in the health score. This main effect is quite large when compared to the much studied education-health gradient. For example, the unconditional mean difference in health between native-born tertiary educated individuals and native-born high school dropouts is slightly less than one half of a standard deviation in the health score. Controlling for an increasingly comprehensive set of control variables has only a small effect on the magnitude of the estimates which makes us confident that the instrument is indeed unrelated to pre-treatment characteristics.[7]

Measurement error in English deficiency

In the likely presence of (non-classical) measurement error in the binary indicator for English deficiency, the OLS estimator will be downward biased. Under the reasonable assumption that, conditional upon the true probability of English deficiency, the measurement errors in both the observed endogenous variable and the instrument are independent of each other and physical health, it is readily shown (Kane et al., 1999) that, even with a valid instrument, the IV estimator is biased by a factor that depends only on the misclassification rates in the endogenous variable.[8]

We use the ALLS data on subjective and objective language proficiency to assess these probabilities. Restricting the sample to be analogous to the HILDA sample, we have 432 individuals with valid information on self reported language skills and an objective assessment of adult literacy skills along a continuous scale from 0 to 500. A

[7]We refrain from using alternative self-reported measures of health in our main specifications due to additionally necessary assumptions on inter-temporal and inter-personal comparability and additional measurement issues. Nonetheless, estimated effects of English deficiency on a five-point scale of self-reported health (-.761[.55]) and a binary indicator of self-reported long-term health impairments (-.289[.17]) yield comparable results. Additional information on doctor visits and hospitalization are not regularly included in the HILDA data.

[8]For example, consider a model with a binary endogenous regressor (English deficiency) and a binary instrument based upon the interaction of later arrivers and non-English mother tongue, both of which are subject to misclassification errors. Let w be a binary variable that takes on a value of one if an individual is truly proficient in English and let $\tilde{w}$ be a binary variable that takes on a value of one if an individual self-reports to be proficient in English. Define the correctly classified observations:

$$\begin{aligned} q_1 &= \Pr(\tilde{w}_i = 1 | w_i = 1) \\ 1 - q_0 &= \Pr(\tilde{w}_i = 0 | w_i = 0) \end{aligned}$$

and the misclassified observations:

$$\begin{aligned} q_0 &= \Pr(\tilde{w}_i = 1 | w_i = 0) \\ 1 - q_1 &= \Pr(\tilde{w}_i = 0 | w_i = 1) \end{aligned}$$

It can be shown that the IV estimator is biased by a factor $1/(q_1 - q_0)$.

skill score of 225 is regarded by the survey developers as the 'minimum required for individuals to meet the complex demands of everyday life and work in the emerging knowledge-based economy' (Statistics Canada, 2005). Accordingly, individuals with a skill score below 225 would be deemed as functionally illiterate. Based upon this threshold, the ALLS sample provides an estimate of 0.220 for this 'true' probability of English deficiency.

To map the continuously measured literacy scores of the ALLS into our binary language deficiency measure, we estimate a Probit model of the binary indicator on the continuous literacy test score and to estimate the rates of misclassification according to different (arbitrarily chosen) cutoffs based upon the predicted probabilities. For example, assigning a cutoff of 0.50 implies that all individuals with a predicted probability less than 0.5 are treated as (truly) English deficient and all individuals with a predicted probability at least as large as 0.50 are treated as (truly) English proficient. In this case, together with an estimate of the 'true' probability of English deficiency, the estimates of the conditional probabilities q_1 and q_0 may be calculated. For a cutoff of 0.5, the OLS estimator is downward biased by an approximate factor of 0.55 while the IV estimator would be upward biased by an approximate factor of 1.610.

Increasing the cutoff threshold to 0.60 and 0.75, implies the IV estimator would be upward biased by a factor of 2.015, or 3.420, respectively. This analysis, based upon the auxiliary ALLS sample, provides suggestive evidence that our finding that the IV estimates are almost twice as large as the OLS estimates is consistent with the presence of non-classical measurement error in self-reported language skills. Consequently, the OLS results in the Table 5 provide a lower bound on the effect of language deficiency on health and the IV estimates provide an upper bound.

5 Conclusion

The acquisition of the destination language is crucial for the social and economic integration of migrants into the destination country, and its value in the labor market has been extensively analyzed in terms of earnings and employment probabilities. In this paper, we analyzed the importance of language skills on the health outcomes of immigrants.

To identify the causal effect of language on the health outcomes, we utilized a structural break in the ability of acquiring new languages during childhood. Information on linguistic dissimilarity between home and destination language and on age at arrival allows us to construct an instrumental variable for language skills that resembles a difference-in-difference estimator comparing immigrants from different linguistic backgrounds arriving at younger and older ages.

Our estimates imply a large and distinctive negative causal effect of language deficiency on the physical health score of immigrants. This strong impact of language on health might be explained within the theoretical framework of a Grossman health production model: Language skills play a twofold role in affecting the health acquisition of immigrants, shaping the access to inputs into the health production (such as employment, occupational choice, wages, and social integration), but also enable individuals to more efficiently use these inputs for maintaining their health status.

The magnitude of the negative health-language gradient highlights the presence of considerable non-market returns to language skills and provides evidence for the importance of early supportive interventions for immigrants such as language classes and the provision of interpreter services to reduce the language barriers in health care.

Acknowledgements & Conflict of Interest

We thank Mathias Sinning for his initial input and contribution, the participants of ESPE 2014 in Braga, the SOLE 2015 in Montreal, and seminars in Jena and Hannover for helpful comments. This paper uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Project was initiated and is funded by the Australian Government Department of Social Services (DSS) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this paper, however, are those of the author and should not be attributed to either DSS or the Melbourne Institute. No potential conflicts by financial or personal relationships exist that might bias our work. This work was not subject to any ethical committee approval.

References

Aoki, Y. and Santiago, L. (2015). Education, Health and Fertility of UK Immigrants: The Role of English Language Skills. IZA Discussion Papers 9498, Institute for the Study of Labor (IZA).

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 13: 169–181.

Bleakley, H. and Chin, A. (2004). Language skills and earnings: Evidence from childhood immigrants. The Review of Economics and Statistics 86: 481–496.

Brachet, T. (2008). Maternal smoking, misclassification, and infant health. MPRA Paper 21466, University Library of Munich, Germany.

Charette, M. and Meng, R. (1994). Explaining language proficiency: Objective versus self-assessed measures of literacy. Economics Letters 44: 313–321.

Chiswick, B. R. and Miller, P. W. (1995). The endogeneity between language and earnings: International analyses. Journal of Labor Economics 13: 246–288.

Clarke, A. (2015). Age at immigration and the educational attainment of foreign-born children in the United States: The confouding effects of parental education. Working paper, Department of Economics, University of Melbourne.

Coulon, A. de and Wolff, F.-C. (2007). Language proficiency of immigrants and misclassification. Applied Economics Letters 14: 857–861.

Cutler, D. M. and Lleras-Muney, A. (2006). Education and Health: Evaluating Theories and Evidence. NBER Working Paper 12352, National Bureau of Economic Research.

De Walt, D. A., Berkman, N. D., Shreidan, S., Lohr, K. and Pignone, M. P. (2004).

Literacy and health outcomes: A systematic review of the literature. Journal of General Internal Medicine 12: 1228–3639.

Dustmann, C. and van Soest, A. (2001). Language fluency and earnings: Estimation with misclassified language indicators. The Review of Economics and Statistics 83: 663–674.

Dustmann, C. and van Soest, A. (2002). Language and the earnings of immigrants. Industrial and Labor Relations Review 55: 473–492.

Eichler, K., Wieser, S. and Brügger, U. (2009). The costs of limited health literacy: a systematic review. International Journal of Public Health 54: 313–324.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. Journal of Political Economy 80: 223–55.

Grossman, M. and Kaestner, R. (1997). Effects of education on health. In Behrman, J. and Stacey, N. (eds), The Social Benefits of Education. University of Michigan Press, 69–123.

Guven, C. and Islam, A. (2015). Age at migration, language proficiency and socioeconomic outcomes: Evidence from australia. Demography 52: 513–542.

Inglehart, R. and Welzel, C. (2005). Modernization, Cultural Change and Democracy. New York: Cambridge University Press.

Isphording, I. E. (2014). Disadvantages of linguistic origin - Evidence from immigrant literacy scores. Economics Letters 123: 236–239.

Isphording, I. E. and Otten, S. (2013). The costs of babylon - linguistic distance in applied economics. Review of International Economics 21: 354–369.

Kane, T. J., Rouse, C. E. and Staiger, D. (1999). Estimating returns to schoolong when schooling is misreported. NBER Working Paper 7235, National Bureau of Economic Research.

Lieberson, S. and Dil, A. S. (1981). Language diversity and language contact: essays, 16. Stanford University Press.

McLaughlin, B. (1992). Myths and Miscoceptions about Second Language Learning: What Every Teachers Needs to Unlearn. Educational Practice Report 7, National Center for Research on Cultural Diversity and Second Language Learning.

Newport, E. (2002). Critical periods in language development. In Nadel, L. (ed.), Encyclopedia of Cognitive Science. London: Macmillan Publishers Ltd./Nature Publishing Group.

Spolaore, E. and Wacziarg, R. (2009). The diffusion of development. The Quarterly Journal of Economics 124: 469–529.

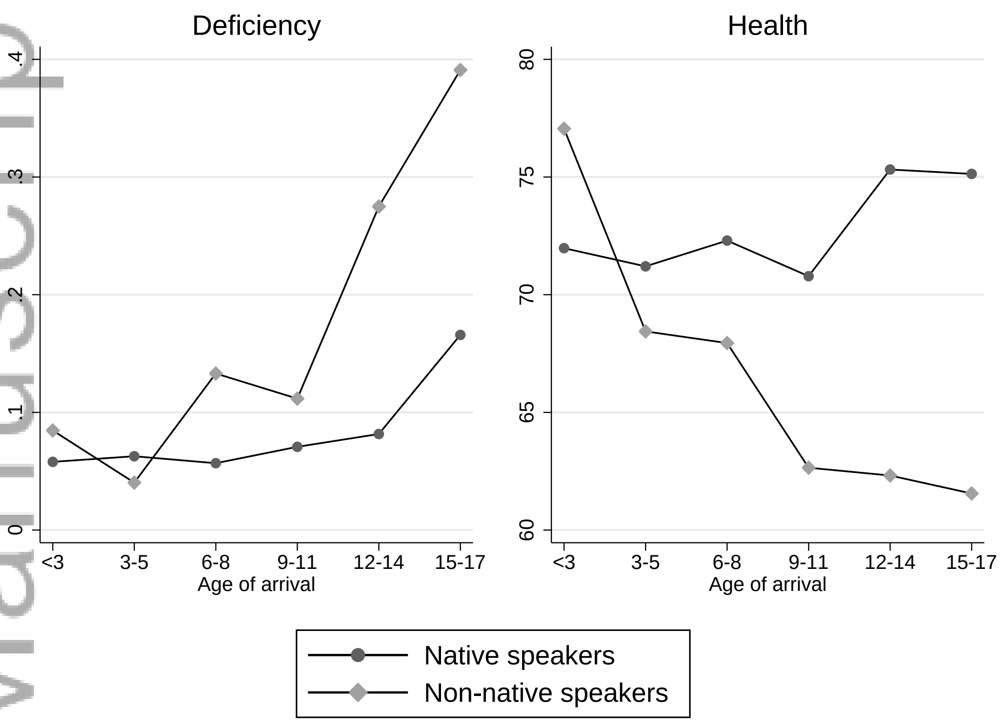
Statistics Canada (2005). Learning a Living: First Results of the Adult Literacy and Life Skills Survey. Minister of Industry, Canada and, Organisation for Economic Cooperation and Development (OECD), Paris.

van den Berg, G. J., Lundborg, P., Nystedt, P. and Rooth, D.-O. (2014). Critical periods during childhood and adolescence. Journal of the European Economic Association 12: 1521–1557.

Wichmann, S., Holman, E. W., Bakker, D. and Brown, C. H. (2010). Evaluating linguistic distance measures. Physica A: Statistical Mechanics and its Applications 389: 3632–3639.

Tables and Figures

Figure 1: Mean English deficiency and physical health by age at arrival



Notes: This figure shows mean language deficiency and SF-36 physical health scores by age at arrival cohort, conditioning on a set of covariates: gender, age (in categories), parental education, urban environment, state indicators and SEIFA-IRSAD as well as time and country of origin fixed effects.

Table 1: The ASJP Linguistic distance measure

I	You	We	One
Two	Person	Fish	Dog
Louse	Tree	Leaf	Skin
Blood	Bone	Horn	Ear
Eye	Nose	Tooth	Tongue
Knee	Hand	Breast	Liver
Drink	See	Hear	Die
Come	Sun	Star	Water
Stone	Fire	Path	Mountain
Night	Full	New	Name

Word	English	German	Distance
fish	*fiS*	fiS	0
breast	*brest*	brust	1
hand	*hEnd*	hant	2
tree	*tri*	baum	4
Mountain	*maunt3n*	bErk	7

Source: Isphording and Otten (2013).

Table 2: Descriptive statistics by linguistic origin and age at arrival

	Total	Native speakers Arrived before 12	Native speakers Arrived after 12	Non-native speakers Arrived before 12	Non-native speakers Arrived after 12
Physical health (SF-36)	78.21	77.56	79.78	80.30	76.68
	(19.85)	(20.44)	(17.90)	(18.19)	(21.03)
English deficiency	0.04	0.00	0.01	0.07	0.35
	(0.16)	(0.01)	(0.10)	(0.16)	(0.36)
Linguistic distance	0.21	0.00	0.00	0.94	0.96
	(0.40)	(0.00)	(0.00)	(0.11)	(0.09)
Subjective childhood health status	3.32	3.34	3.36	3.28	3.25
	(0.98)	(0.97)	(1.03)	(1.00)	(0.90)
Female	0.53	0.49	0.51	0.66	0.59
	(0.50)	(0.50)	(0.50)	(0.47)	(0.49)
Age	41.40	42.50	45.32	34.95	37.06
	(12.36)	(11.31)	(12.37)	(13.41)	(12.84)
Age at arrival	7.44	4.97	14.28	6.47	15.17
	(5.08)	(3.49)	(1.96)	(3.21)	(1.75)
Married	0.70	0.75	0.72	0.53	0.55
	(0.46)	(0.43)	(0.45)	(0.50)	(0.50)
No. of children in household	0.64	0.68	0.46	0.68	0.62
	(1.02)	(1.03)	(0.80)	(1.17)	(0.97)
Household income (in 1,000 AUD)	93.52	93.93	86.47	104.80	84.03
	(83.38)	(83.88)	(62.26)	(107.59)	(60.60)
Full-time employed	0.53	0.55	0.55	0.45	0.48
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Part-time employed	0.23	0.22	0.25	0.26	0.16
	(0.42)	(0.41)	(0.43)	(0.44)	(0.37)
Unemployed	0.03	0.03	0.03	0.04	0.05
	(0.18)	(0.18)	(0.16)	(0.19)	(0.22)
Not in labor force	0.21	0.20	0.18	0.25	0.31
	(0.41)	(0.40)	(0.38)	(0.43)	(0.46)
Education: Tertiary	0.30	0.30	0.21	0.33	0.32
	(0.46)	(0.46)	(0.41)	(0.47)	(0.47)
Education: Certificate/Diploma	0.30	0.31	0.42	0.19	0.17
	(0.46)	(0.46)	(0.49)	(0.39)	(0.38)
Education: Year 12	0.17	0.15	0.11	0.25	0.23
	(0.37)	(0.36)	(0.32)	(0.43)	(0.42)
Education: Year 11 or below	0.24	0.23	0.27	0.23	0.27
	(0.43)	(0.42)	(0.44)	(0.42)	(0.44)
Father has higher education	0.41	0.44	0.35	0.40	0.36
	(0.49)	(0.50)	(0.48)	(0.49)	(0.48)
Mother has higher education	0.27	0.26	0.30	0.32	0.24
	(0.45)	(0.44)	(0.46)	(0.47)	(0.43)
Living in urban area	0.68	0.62	0.69	0.84	0.83
	(0.47)	(0.49)	(0.46)	(0.37)	(0.38)
SEIFA index of relative economic advantages and disadvantages	5.68	5.88	5.62	5.42	4.78
	(2.88)	(2.78)	(2.77)	(3.13)	(3.20)
Number of observations	5,706	3553	867	473	813

Notes: Standard deviations are reported in parentheses

Table 3: Age-at-arrival patterns in physical health

	(1)	(2)	(3)	(4)
Panel A:				
Age at arrival (AAA)				
Ref. cat.: Age at arrival 0-2				
Age at arrival 3-5	-1.519	-1.155	-0.581	0.796
	(2.12)	(2.06)	(2.17)	(2.05)
Age at arrival 6-8	-0.210	-0.144	0.849	1.594
	(2.04)	(2.04)	(2.19)	(2.14)
Age at arrival 9-11	-3.368	-3.036	-2.008	-1.198
	(2.55)	(2.50)	(2.62)	(2.53)
Age at arrival 12-14	0.821	1.573	2.696	4.025
	(2.82)	(2.81)	(2.87)	(2.85)
Age at arrival 15-17	-0.944	0.038	1.007	2.221
	(3.18)	(3.19)	(3.22)	(3.07)
Basic controls	yes	yes	yes	yes
Origin-fixed effects	no	yes	yes	yes
Regional effects	no	no	yes	yes
Socio-economic controls	no	no	no	yes
R^2	0.083	0.103	0.141	0.212
N	5706	5706	5706	5706
Panel B:				
Age at arrival (AAA) by linguistic distance (LD)				
Ref. cat.: Native speaker, Age at arrival 0-2				
Age at arrival 3-5	-1.356	-0.773	0.302	1.681
	(2.36)	(2.27)	(2.36)	(2.22)
Age at arrival 6-8	0.157	0.308	1.576	2.560
	(2.25)	(2.23)	(2.40)	(2.30)
Age at arrival 9-11	-1.651	-1.356	-0.139	0.557
	(2.81)	(2.73)	(2.81)	(2.69)
Age at arrival 12-14	2.492	3.346	4.817	5.903**
	(2.98)	(2.96)	(3.02)	(2.99)
Age at arrival 15-17	2.060	2.725	4.453	5.559*
	(3.56)	(3.54)	(3.61)	(3.36)
Linguistic distance	4.696*	5.133**	3.752	3.436
	(2.45)	(2.57)	(4.32)	(4.21)
LD × AAA 3-5	-2.217	-3.799	-7.059	-6.987
	(4.23)	(4.10)	(5.16)	(4.83)
LD × AAA 6-8	-3.500	-4.204	-5.939	-7.432*
	(3.33)	(3.37)	(4.02)	(3.92)
LD × AAA 9-11	-9.072**	-9.271**	-10.755**	-10.434**
	(4.34)	(4.33)	(4.82)	(4.44)
LD × AAA 12-14	-9.303**	-10.258**	-12.785***	-11.442**
	(4.10)	(4.06)	(4.80)	(4.74)
LD × AAA 15-17	-10.077**	-9.938**	-13.450***	-13.241***
	(4.22)	(4.20)	(4.81)	(4.32)
Basic controls	yes	yes	yes	yes
Origin-fixed effects	no	yes	yes	yes
Regional effects	no	no	yes	yes
Socio-economic controls	no	no	no	yes
R^2	0.089	0.109	0.148	0.218
N	5706	5706	5706	5706

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, reported in parentheses, are clustered at the individual level. The sample is restricted to immigrants aged 18 - 65. The dependent variable is based on SF-36 health scores which range from 0-100. Basic controls include gender, age (in categories), parental education, and time-fixed effects. Regional controls include living in an urban environment, state indicators and SEIFA-IRSAD. Socio-economic controls include marital status, individual education, number of children in household, household income, labor force status.

Table 4: Age-at-arrival patterns in English deficiency

	(1)	(2)	(3)	(4)
Panel A: Age at arrival (linear) by linguistic distance				
Age at arrival	0.003**	0.002**	0.002	0.002
	(0.00)	(0.00)	(0.00)	(0.00)
Non-native speaker	-0.086***	-0.090***	-0.158***	-0.154***
	(0.03)	(0.03)	(0.04)	(0.04)
Interaction	0.027***	0.027***	0.027***	0.027***
	(0.00)	(0.00)	(0.00)	(0.00)
Basic controls	yes	yes	yes	yes
Origin-fixed effects	no	yes	yes	yes
Regional effects	no	no	yes	yes
Socio-economic controls	no	no	no	yes
R^2	0.373	0.383	0.536	0.546
N	5706	5706	5706	5706
Panel B: Age at arrival (AAA) by linguistic distance (LD) Ref. cat.: Native speaker, age at arrival 0-2				
Age at arrival 3-5	0.003	0.002	0.004	0.002
	(0.00)	(0.00)	(0.01)	(0.01)
Age at arrival 6-8	0.000	-0.002	0.000	0.000
	(0.01)	(0.01)	(0.01)	(0.01)
Age at arrival 9-11	0.008	0.006	0.009	0.009
	(0.01)	(0.01)	(0.01)	(0.01)
Age at arrival 12-14	0.013	0.010	0.011	0.007
	(0.01)	(0.01)	(0.01)	(0.01)
Age at arrival 15-17	0.057**	0.056**	0.049**	0.045**
	(0.03)	(0.02)	(0.02)	(0.02)
Linguistic distance	0.041	0.030	-0.051	-0.048
	(0.03)	(0.03)	(0.04)	(0.04)
LD × AAA 3-5	-0.028	-0.018	-0.003	-0.005
	(0.03)	(0.03)	(0.04)	(0.04)
LD × AAA 6-8	0.079*	0.090*	0.120**	0.121**
	(0.05)	(0.05)	(0.06)	(0.05)
LD × AAA 9-11	0.056	0.064	0.067	0.062
	(0.04)	(0.04)	(0.04)	(0.04)
LD × AAA 12-14	0.233***	0.239***	0.245***	0.239***
	(0.09)	(0.09)	(0.08)	(0.08)
LD × AAA 15-17	0.345***	0.347***	0.370***	0.369***
	(0.06)	(0.06)	(0.06)	(0.06)
Basic controls	yes	yes	yes	yes
Origin-fixed effects	no	yes	yes	yes
Regional effects	no	no	yes	yes
Socio-economic controls	no	no	no	yes
R^2	0.408	0.416	0.568	0.580
N	5706	5706	5706	5706

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, reported in parentheses, are clustered at the individual level. The sample is restricted to immigrants aged 18 - 65. The dependent variable is English deficiency coded 1 if individuals report not to speak English "Very Well". Basic controls include gender, age (in categories), parental education, and time- and cohort fixed effects. Regional controls include living in an urban environment, state indicators and SEIFA-IRSAD. Socio-economic controls include marital status, individual education, number of children in household, household income, labor force status.

Table 5: IV results: The effect of English deficiency on physical health

	English Deficiency	Physical Health				
	OLS	OLS	IV 1	IV 2	IV 3	IV 4
English deficiency	–	-11.702***	-20.506*	-19.658*	-23.736**	-20.427**
		(3.83)	(10.51)	(10.50)	(10.94)	(10.29)
LD * (AAA > 11)	0.269***	–	–	–	–	–
	(0.04)					
Ref. cat.: Age at arrival 0-2						
Age at arrival 3-5	-0.000	1.061	-1.566	-1.161	-0.292	1.044
	(0.01)	(2.03)	(2.10)	(2.04)	(2.13)	(2.01)
Age at arrival 6-8	0.019*	2.103	0.092	0.135	1.511	2.203
	(0.01)	(2.13)	(2.03)	(2.03)	(2.17)	(2.11)
Age at arrival 9-11	0.016	-0.684	-3.028	-2.722	-1.412	-0.643
	(0.01)	(2.49)	(2.50)	(2.44)	(2.53)	(2.45)
Age at arrival 12-14	-0.004	4.985*	2.080	2.733	4.240	5.338*
	(0.01)	(2.82)	(2.86)	(2.83)	(2.87)	(2.82)
Age at arrival 14-17	0.070***	4.667	3.000	3.756	5.397	6.025*
	(0.03)	(3.12)	(3.81)	(3.79)	(3.82)	(3.57)
Linguistic distance	-0.006	-4.032	2.018	1.690	-2.765	-3.207
	(0.03)	(3.11)	(2.10)	(2.06)	(3.09)	(3.10)
Basic controls	yes	yes	yes	yes	yes	yes
Origin-fixed effects	yes	yes	no	yes	yes	yes
Regional effects	yes	yes	no	no	yes	yes
Socio-economic controls	yes	yes	no	no	no	yes
R^2	0.57	0.22	0.10	0.11	0.15	0.16
F-test of excluded instruments			30.85	30.65	33.99	35.88
N	5706	5706	5706	5706	5706	5706

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, reported in parentheses, are clustered at the individual level. The sample restrictions are identical to those for Table 3. The dependent variables is based on SF-36 health scores which range from 0-100. Basic controls include gender, age (in categories), parental education, and time- and cohort fixed effects. Regional controls include living in an urban environment, state indicators and SEIFA-IRSAD. Socio-economic controls include marital status, individual education, number of children in household, household income, labor force status.

6 Appendix

Robustness checks

The IV estimates presented in section 4 rely on the exclusion restriction that the interaction between linguistic origin and age at immigration increases the costs of language acquisition yet has no further direct or indirect effect on health production (conditional on our comprehensive controls). Table A.2 summarizes a series of robustness checks that test potential threats to this identifying assumption.

Childhood health and infant mortality

We have already stressed the importance of controlling for initial health status, which we do by including self-reported childhood health in every specification. One might worry that childhood health, as a self-reported variable, suffers from significant recall error and might not be a good proxy for the true initial status. In light of this, we run a robustness check using mean infant mortality in the origin country interacted with age at arrival as an alternative proxy measure. An examination of Table column (2) reveals that this alternative specification have only a negligible influence on the estimated coefficient of English deficiency.

Cultural differences

Linguistic distance is an important part of the socio-cultural differences between countries. Certainly, it is also related to further additional dimensions of cultural, political or institutional differences which are difficult to measure and omitted in our main specification. Such omitted differences do not display a threat to identification per se. However, they represent a possible threat to identification if cultural differences directly affect health through mechanisms apart from through language acquisition <u>and</u> these cultural differences are correlated with the interaction between age at migration and linguistic distance. As we cannot rule out these potential threats of identification, we re-estimate our preferred specification adding interactions between age at arrival and proxy variables for geographical and cultural differences. *Geographic distance* is calculated as the simple geodesic (great circle) distance between source country capital cities and Canberra, the national capital of Australia.

Cultural differences are difficult to map into a continous variable, and different attempts in the literature suffer from different kinds of advantages and disadvantages. We rely on two different proxy variables. The first proxy variable utilizes the popular data provided by Spolaore and Wacziarg (2009) based on genetic similarities across countries and ethnicities. Similarities in frequencies of *alleles* (representations of certain gene combinations), across populations, has been used as a summary statistic for divergence in the *'whole set of implicit beliefs, customs, habits, biases, conventions, etc. that are transmitted across generations – biologically and/or culturally – with high persistence'*(Spolaore and Wacziarg, 2009, p. 471). The idea is that populations, once divided by migration and geographical barriers, start to gradually diverge in their genome, but also in their habits and customs. As such, divergence in distributions of gene variants can be interpreted as a proxy for cultural distance.

The second proxy for cultural differences relies on the work by Inglehart and Welzel (2005) who use answers to the *World Value Survey* to map surveyed populations on a two-dimensional map consisting of dimensions of 'Traditional vs. Secular-Rational Values' and 'Survival vs. Self-Expression Values'. While relying on actual and recent answers to a a large set of survey questions on attitudes, values and norms, it comes at the expense of a lower worldwide coverage and is only available for about 70% of our sample.

Despite very different underlying concepts, both measures are strongly correlated ($r = 0.69$). The results in column (3) and (4) are comparable in magnitude with the baseline results, though become insignificant in the case of using the cultural distance based on the World Value Survey, likely due to a lower sample size.

Excluding Asian countries of origin

Table A.1 indicates that the group of immigrants with a foreign mother tongue originate from a large variety of primarily European and Asian countries. The large share of Asian immigrants, mainly as a result of their geographical proximity to Australia, raises concerns that the results are driven by some systematic (unobserved) characteristic of this group. Estimating our preferred specification excluding the group of (East) Asian immigrants does not support this conjecture (column 5). The estimated effect remains virtually unchanged but becomes insignificant with a decreased sample size.

Excluding multilingual countries of origin

The HILDA dataset only provides information on whether English is the mother tongue or not. Without knowing the exact non-English mother tongue, we assign languages by country of birth, based on the predominant language of a country. This potentially induces measurement error in the linguistic distance variable and in its interaction with age at migration. We reasonably expect this measurement error to be uncorrelated with the measurement error in the English deficiency. In this case, misclassified linguistic distance due to unknown true mother tongue does not necessarily impose an econometric problem.

Nonetheless, in column (6) we report results for the estimation excluding multilingual countries. Multilingual countries are identified using a linguistic diversity index defined by Lieberson and Dil (1981) and based on recent language family population data taken from Ethnologue. This index is defined as the theoretical probability that two random individuals from a country speak different mother tongues. We define countries as multilingual if this index is 0.5 or larger. By excluding these countries from the sample, we are losing 1.146 individual observations from 34 home countries (New Caledonia, Papua New Guinea, Solomon Islands, Marshall Islands, Nauru, Fiji, Belgium, Switzerland, Albania, Bosnia and Herzegovina, Former Yugoslav Republic of Macedonia, Former Yugoslavia, Libya, Bahrain, Iran, Iraq, Oman, Syria, United Arab Emirates, Burma (Myanmar), Laos, Thailand, Brunei Darussalam, Indonesia, Malaysia, Philippines, Singapore, East Timor, China, India, Nepal, Pakistan, Afghanistan, Canada, Trinidad and Tobago, Sierra Leone, Ethiopia, Kenya, Malawi, Mauritius, Mozambique, South Africa, Tanzania, Zambia, and Zimbabwe). The estimated coefficient increases moderately compared to the baseline level.

Between estimator

While several of the variables in our preferred specification vary both across and within individuals (such as English language deficiency, health status, and other controls variables), our identifying instrument only varies across individuals. In order to investigate whether this has any impact upon our results, we estimate equation (3) after collapsing our data to within-person means of English deficiency and health scores, and controlling for only time-invariant characteristics (gender, parental education, country of origin, and childhood health). This effectively ignores any additional within-person variation in language deficiency, health status and the control variables. The point estimate in column (7) exceeds the baseline level, but remains close to the previous results. We conclude from this that using repeated individual observations does not severely alter the observed pattern.

Different SF-36 domains

To further test the robustness of the results and to analyze whether certain subdomains of the SF36 score are driving the results we ran our preferred specification using the scores for the different sub-domains of the SF-36 as dependent variables. Table A.3 provides information on these sub-domains. The results are summarized in Table A.4. Estimates differ moderately in magnitude, but the results do not indicate a clear 'leading' domain—there is no evidence that one single sub-domain is driving our main result. Largest effects are found for the domain of 'physical pain' and lowest effects for 'General health'.

Table A.1: Sample composition by country of origin

Country of origin	Obs.	Share of English as first language	Share of English Deficiency
United Kingdom	2646	1.00	0.00
New Zealand	533	1.00	0.00
Vietnam	176	0.00	0.20
South Africa	157	1.00	0.00
Philippines	146	0.33	0.10
Papua New Guinea	131	1.00	0.00
Italy	119	0.18	0.13
India	100	0.73	0.03
United States of America	98	1.00	0.00
Netherlands	83	0.48	0.02
Canada	79	1.00	0.00
Germany	78	0.10	0.05
Hong Kong (SAR of China)	77	0.29	0.12
Sri Lanka	70	0.74	-0.00
Fiji	69	0.17	0.26
China (excludes SARs and Taiwan)	67	0.00	0.25
Ireland	62	1.00	0.00
Lebanon	54	0.19	0.24
Malta	52	1.00	0.13
Yugoslavia Federal Republic of	46	0.09	0.07
Zimbabwe	44	1.00	0.00
Egypt	43	0.26	0.14
Taiwan	40	0.22	0.22
Malaysia	38	0.39	0.13
Greece	35	0.00	0.46
Other	663	0.46	0.10

Notes: Table lists the top 25 countries ranked by number of person-year observations, with the share of native speakers and English-deficient speakers by country of birth. See notes to Table 3 for the sample restrictions. The residual category "Other" consists of 53 distinct countries of origin with less than 35 observations per country.

Table A.2: Robustness checks

	(1) Baseline	(2) AAA × infant mortality	(3) AAA × geogr. dist & gen. dist	(4) AAA × geogr. dist & WVS dist.	(5) w/o Asian countries	(6) w/o multi-lingual countries	(7) Between-estimator
English deficiency	-20.427**	-19.882*	-17.277*	-22.190	-19.287	-24.627**	-26.013**
	(10.29)	(11.06)	(9.96)	(13.54)	(12.40)	(11.76)	(12.30)
Basic controls	yes	yes	yes	yes	yes	yes	a
Origin-fixed effects	yes	yes	yes	yes	yes	yes	a
Regional effects	yes	yes	yes	yes	yes	yes	a
Socio-economic controls	yes	yes	yes	yes	yes	yes	a
R^2	0.16	0.16	0.17	0.19	0.18	0.18	0.06
F-test of excluded instruments	35.88	35.33	36.95	29.53	24.70	33.47	23.30
N	5706	5543	5491	3946	4560	4773	847

Notes: Robustness checks conducted are: (1) baseline results as in Table 5, column (6), (2) controlling for infant mortality × age at arrival (3) controlling for geographic and genetic distance × age at arrival (4) controlling for geographic and cultural distance (WVS) × age at arrival (5) excluding multilingual countries (6) excluding Asian countries (7) collapsing data to within-person means, (a) controlling only for time-invariant characteristics (gender, parental education, country of origin, childhood health).

Table A.3: SF–36 domains of physical health

Items (34)	Domains (8)	Summary Measures (2)
Vigorous Activities	Physical functioning	**Physical Health**
Moderate Activities		
Lift, Carry, Groceries		
Climb Several Flights		
Climb One Flight		
Bend, Kneel		
Walk Mile		
Walk Several Blocks		
Walk One Block		
Bathe, Dress		
Cut Down Time	Role-Physical	
Accomplished Less		
Limited in Kind		
Had Difficulty		
Pain-Magnitude	Bodily Pain	
Pain-Interfere		
Sick Easier	General Health	
As Healthy as anyone		
Health To Get Worse		
Health Excellent		

Source: www.sf-36.org.

Table A.4: IV results: The effect of English deficiency on physical health by SF-36 domain

	Physical functioning	Role-physical	Bodily pain	General health
Panel A: Baseline specification				
English deficiency	-21.830**	-27.118*	-28.359**	-17.638
	(10.32)	(15.54)	(12.81)	(13.17)
N	5706	5706	5706	5706
Panel B: Including socioeconomic controls				
English deficiency	-19.430**	-19.692	-27.500**	-15.088
	(9.50)	(14.43)	(12.51)	(13.11)
N	5706	5706	5706	5706

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, reported in parentheses, are clustered at the individual level. The sample restrictions are identical to those for Table 3. The dependent variables is based on SF-36 health scores which range from 0-100. Basic controls include gender, age (in categories), parental education, and time- and cohort fixed effects. Regional controls include living in an urban environment, state indicators and SEIFA-IRSAD. Socio-economic controls include marital status, individual education, number of children in household, household income, labor force status.

Web Appendix for Language Barriers and Immigrant Health

Andrew Clarke*and Ingo E. Isphording †

April 1, 2016

Abstract

We study the impact of language deficiency on the health status of childhood migrants to Australia. Our identification strategy relies on a quasi-experiment comparing immigrants arriving at different ages and from different linguistic origins. In the presence of considerable non-classical measurement error in self-reported language proficiency, our results provide lower and upper bounds for a strong negative effect of English deficiency on health of between one half and a full standard deviation in the health score.

Keywords: International Migration; Language Skills; Health.
JEL Classification: F22, I12, J24, J61

*Department of Economics, University of Melbourne, Level 4, Faculty of Business and Economics Building, 111 Barry Street, Carlton, Victoria, Australia; andrew.clarke@unimelb.edu.au.

†Corresponding author: Ingo E. Isphording, Institute for the Study of Labor, Schaumburg-Lippe-Str. 5-9, 53113 Bonn, Germany. Phone, +49 (228) 3894-204, Fax +49 (228) 3894-510, E-mail isphording@iza.org.

Appendix

Robustness checks

The IV estimates presented in section 4 rely on the exclusion restriction that the interaction between linguistic origin and age at immigration increases the costs of language acquisition yet has no further direct or indirect effect on health production (conditional on our comprehensive controls). Table A.2 summarizes a series of robustness checks that test potential threats to this identifying assumption.

Childhood health and infant mortality

We have already stressed the importance of controlling for initial health status, which we do by including self-reported childhood health in every specification. One might worry that childhood health, as a self-reported variable, suffers from significant recall error and might not be a good proxy for the true initial status. In light of this, we run a robustness check using mean infant mortality in the origin country interacted with age at arrival as an alternative proxy measure. An examination of Table column (2) reveals that this alternative specification have only a negligible influence on the estimated coefficient of English deficiency.

Cultural differences

Linguistic distance is an important part of the socio-cultural differences between countries. Certainly, it is also related to further additional dimensions of cultural, political or institutional differences which are difficult to measure and omitted in our main specification. Such omitted differences do not display a threat to identification per se. However, they represent a possible threat to identification if cultural differences directly affect health through mechanisms apart from through language acquisition and these cultural differences are correlated with the interaction between age at migration and linguistic distance. As we cannot rule out these potential threats of identification, we re-estimate our preferred specification adding interactions between age at arrival and proxy variables for geographical and cultural differences. *Geographic distance* is calculated as the simple geodesic (great circle) distance between source country capital cities and Canberra, the national capital of Australia.

Cultural differences are difficult to map into a continous variable, and different attempts in the literature suffer from different kinds of advantages and disadvantages. We rely on two different proxy variables. The first proxy variable utilizes the popular data provided by Spolaore and Wacziarg (2009) based on genetic similarities across countries and ethnicities. Similarities in frequencies of *alleles* (representations of certain gene combinations), across populations, has been used as a summary statistic for divergence in the *'whole set of implicit beliefs, customs, habits, biases, conventions, etc. that are transmitted across generations – biologically and/or culturally – with high persistence'*(Spolaore and Wacziarg, 2009, p. 471). The idea is that populations, once divided by migration and geographical barriers, start to gradually diverge in their genome, but also in their habits and customs. As such, divergence in distributions of gene variants can be interpreted as a proxy for cultural distance.

The second proxy for cultural differences relies on the work by Inglehart and Welzel (2005) who use answers to the *World Value Survey* to map surveyed populations on a two-dimensional map consisting of dimensions of 'Traditional vs. Secular-Rational Values' and 'Survival vs. Self-Expression Values'. While relying on actual and recent answers to a a large set of survey questions on attitudes, values and norms, it comes at the expense of a lower worldwide coverage and is only available for about 70% of our sample.

Despite very different underlying concepts, both measures are strongly correlated ($r = 0.69$). The results in column (3) and (4) are comparable in magnitude with the baseline results, though become insignificant in the case of using the cultural distance based on the World Value Survey, likely due to a lower sample size.

Excluding Asian countries of origin

Table A.1 indicates that the group of immigrants with a foreign mother tongue originate from a large variety of primarily European and Asian countries. The large share of Asian immigrants, mainly as a result of their geographical proximity to Australia, raises concerns that the results are driven by some systematic (unobserved) characteristic of this group. Estimating our preferred specification excluding the group of (East) Asian immigrants does not support this conjecture (column 5). The estimated effect remains virtually unchanged but becomes insignificant with a decreased sample size.

Excluding multilingual countries of origin

The HILDA dataset only provides information on whether English is the mother tongue or not. Without knowing the exact non-English mother tongue, we assign languages by country of birth, based on the predominant language of a country. This potentially induces measurement error in the linguistic distance variable and in its interaction with age at migration. We reasonably expect this measurement error to be uncorrelated with the measurement error in the English deficiency. In this case, misclassified linguistic distance due to unknown true mother tongue does not necessarily impose an econometric problem.

Nonetheless, in column (6) we report results for the estimation excluding multilingual countries. Multilingual countries are identified using a linguistic diversity index defined by Lieberson and Dil (1981) and based on recent language family population data taken from Ethnologue. This index is defined as the theoretical probability that two random individuals from a country speak different mother tongues. We define countries as multilingual if this index is 0.5 or larger. By excluding these countries from the sample, we are losing 1.146 individual observations from 34 home countries (New Caledonia, Papua New Guinea, Solomon Islands, Marshall Islands, Nauru, Fiji, Belgium, Switzerland, Albania, Bosnia and Herzegovina, Former Yugoslav Republic of Macedonia, Former Yugoslavia, Libya, Bahrain, Iran, Iraq, Oman, Syria, United Arab Emirates, Burma (Myanmar), Laos, Thailand, Brunei Darussalam, Indonesia, Malaysia, Philippines, Singapore, East Timor, China, India, Nepal, Pakistan, Afghanistan, Canada, Trinidad and Tobago, Sierra Leone, Ethiopia, Kenya, Malawi, Mauritius, Mozambique, South Africa, Tanzania, Zambia, and Zimbabwe). The estimated coefficient increases moderately compared to the baseline level.

Between estimator

While several of the variables in our preferred specification vary both across and within individuals (such as English language deficiency, health status, and other controls variables), our identifying instrument only varies across individuals. In order to investigate whether this has any impact upon our results, we estimate equation (3) after collapsing our data to within-person means of English deficiency and health scores, and controlling for only time-invariant characteristics (gender, parental education, country of origin, and childhood health). This effectively ignores any additional within-person variation in language deficiency, health status and the control variables. The point estimate in column (7) exceeds the baseline level, but remains close to the previous results. We conclude from this that using repeated individual observations does not severely alter the observed pattern.

Different SF-36 domains

To further test the robustness of the results and to analyze whether certain subdomains of the SF36 score are driving the results we ran our preferred specification using the scores for the different sub-domains of the SF-36 as dependent variables. Table A.3 provides information on these sub-domains. The results are summarized in Table A.4. Estimates differ moderately in magnitude, but the results do not indicate a clear 'leading' domain—there is no evidence that one single sub-domain is driving our main result. Largest effects are found for the domain of 'physical pain' and lowest effects for 'General health'.

Table A.1: Sample composition by country of origin

Country of origin	Obs.	Share of English as first language	Share of English Deficiency
United Kingdom	2646	1.00	0.00
New Zealand	533	1.00	0.00
Vietnam	176	0.00	0.20
South Africa	157	1.00	0.00
Philippines	146	0.33	0.10
Papua New Guinea	131	1.00	0.00
Italy	119	0.18	0.13
India	100	0.73	0.03
United States of America	98	1.00	0.00
Netherlands	83	0.48	0.02
Canada	79	1.00	0.00
Germany	78	0.10	0.05
Hong Kong (SAR of China)	77	0.29	0.12
Sri Lanka	70	0.74	-0.00
Fiji	69	0.17	0.26
China (excludes SARs and Taiwan)	67	0.00	0.25
Ireland	62	1.00	0.00
Lebanon	54	0.19	0.24
Malta	52	1.00	0.13
Yugoslavia Federal Republic of	46	0.09	0.07
Zimbabwe	44	1.00	0.00
Egypt	43	0.26	0.14
Taiwan	40	0.22	0.22
Malaysia	38	0.39	0.13
Greece	35	0.00	0.46
Other	663	0.46	0.10

Notes: Table lists the top 25 countries ranked by number of person-year observations, with the share of native speakers and English-deficient speakers by country of birth. See notes to Table 3 for the sample restrictions. The residual category "Other" consists of 53 distinct countries of origin with less than 35 observations per country.

Table A.2: Robustness checks

	(1) Baseline	(2) AAA × infant mortality	(3) AAA × geogr. dist & gen. dist	(4) AAA × geogr. dist & WVS dist.	(5) w/o Asian countries	(6) w/o multi-lingual countries	(7) Between-estimator
English deficiency	-20.427**	-19.882*	-17.277*	-22.190	-19.287	-24.627**	-26.013**
	(10.29)	(11.06)	(9.96)	(13.54)	(12.40)	(11.76)	(12.30)
Basic controls	yes	yes	yes	yes	yes	yes	a
Origin-fixed effects	yes	yes	yes	yes	yes	yes	a
Regional effects	yes	yes	yes	yes	yes	yes	a
Socio-economic controls	yes	yes	yes	yes	yes	yes	a
R^2	0.16	0.16	0.17	0.19	0.18	0.18	0.06
F-test of excluded instruments	35.88	35.33	36.95	29.53	24.70	33.47	23.30
N	5706	5543	5491	3946	4560	4773	847

Notes: Robustness checks conducted are: (1) baseline results as in Table 5, column (6), (2) controlling for infant mortality × age at arrival (3) controlling for geographic and genetic distance × age at arrival (4) controlling for geographic and cultural distance (WVS) × age at arrival (5) excluding multilingual countries (6) excluding Asian countries (7) collapsing data to within-person means, (a) controlling only for time-invariant characteristics (gender, parental education, country of origin, childhood health).

References

Inglehart, R. and Welzel, C. (2005). Modernization, Cultural Change and Democracy. New York: Cambridge University Press.

Lieberson, S. and Dil, A. S. (1981). Language diversity and language contact: essays, 16. Stanford University Press.

Table A.3: SF–36 domains of physical health

Items (34)	Domains (8)	Summary Measures (2)
Vigorous Activities Moderate Activities Lift, Carry, Groceries Climb Several Flights Climb One Flight Bend, Kneel Walk Mile Walk Several Blocks Walk One Block Bathe, Dress	Physical functioning	**Physical Health**
Cut Down Time Accomplished Less Limited in Kind Had Difficulty	Role-Physical	
Pain-Magnitude Pain-Interfere	Bodily Pain	
Sick Easier As Healthy as anyone Health To Get Worse Health Excellent	General Health	

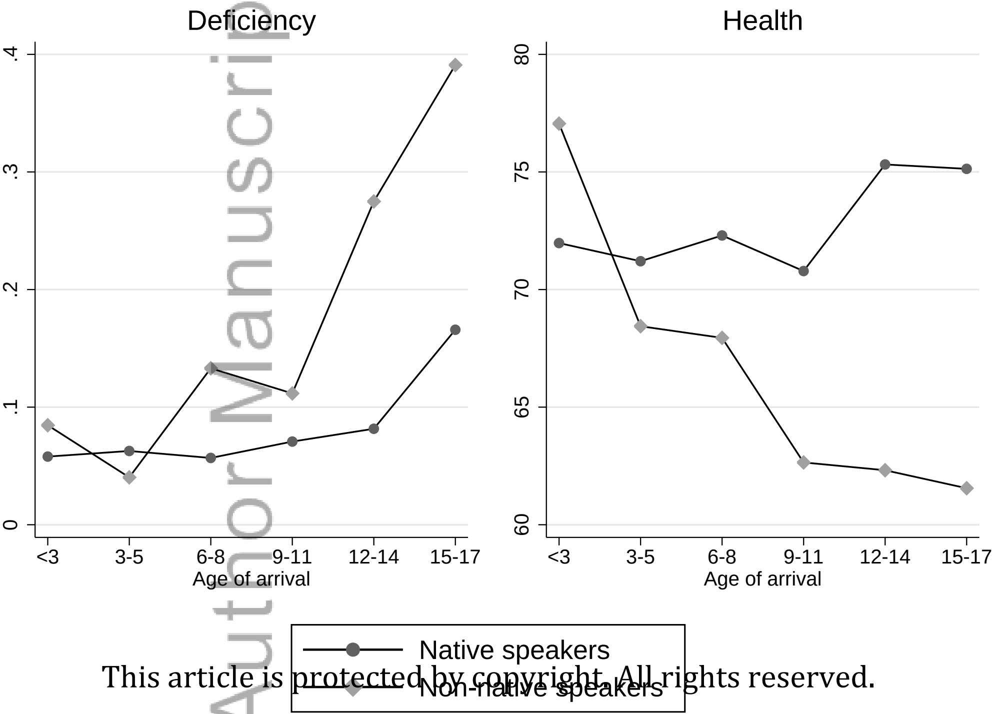
Source: www.sf-36.org.

Table A.4: IV results: The effect of English deficiency on physical health by SF-36 domain

	Physical functioning	Role-physical	Bodily pain	General health
Panel A: Baseline specification				
English deficiency	-21.830**	-27.118*	-28.359**	-17.638
	(10.32)	(15.54)	(12.81)	(13.17)
N	5706	5706	5706	5706
Panel B: Including socioeconomic controls				
English deficiency	-19.430**	-19.692	-27.500**	-15.088
	(9.50)	(14.43)	(12.51)	(13.11)
N	5706	5706	5706	5706

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, reported in parentheses, are clustered at the individual level. The sample restrictions are identical to those for Table 3. The dependent variables is based on SF-36 health scores which range from 0-100. Basic controls include gender, age (in categories), parental education, and time- and cohort fixed effects. Regional controls include living in an urban environment, state indicators and SEIFA-IRSAD. Socio-economic controls include marital status, individual education, number of children in household, household income, labor force status.

Spolaore, E. and Wacziarg, R. (2009). The diffusion of development. The Quarterly Journal of Economics 124: 469–529.


Deficiency
Health
.4
.3
.2
.1
0
80
75
70
65
60
<3
3-5
6-8
9-11
12-14
15-17
Age of arrival
Native speakers
Non-native speakers