CINCH
competence in competition + health
series

#2015/10

Yu Aoki and Lualhati Santiago

Fertility, Health and Education of UK Immigrants:

The Role of English Language Skills

Imprint

CINCH SERIES

CINCH – Health Economics Research Center
Edmund-Körner-Platz 2
45127 Essen

Phone +49 (0) 201 183 - 6326
Fax +49 (0) 201 183 - 3716
Email: daniel.avdic@uni-due.de
Web: www.cinch.uni-due.de

competence in competition + health
series

#2015/10

Yu Aoki and Lualhati Santiago

Fertility, Health and Education of UK Immigrants:

The Role of English Language Skills

UNIVERSITÄT DUISBURG ESSEN
Open-Minded

Bundesministerium für Bildung und Forschung

RWI

Yu Aoki* and Lualhati Santiago

Fertility, Health and Education of UK Immigrants:

The Role of English Language Skills

Abstract

This paper aims to identify the causal effect of English language skills on fertility, health and education outcomes of immigrants in England and Wales. To estimate this causal effect, we use the instrumental variable estimation strategy where age at arrival in the United Kingdom (UK) is exploited to construct an instrument for language skills. The idea of exploiting age at arrival is based on the phenomenon that a person who is exposed to a new language within the critical period of language acquisition (i.e., childhood) learns the language more easily. This implies that immigrants who arrive in the UK at a young age have on average better English language skills than those who arrive when they are older. Using a unique individual-level dataset that links census and life event records for the population living in England and Wales at the 2011 Census, we find that better English language skills significantly delay the age at which a woman has her first child, lower the likelihood that she has a child in her teens, and decrease the number of children she gives birth to, but do not affect her children's birthweight and an individual's self-reported health. The impact on educational achievement is also considerable: better English skills significantly raise the probability of obtaining academic degrees and significantly lower the probability of having no qualifications.

JEL Classifications: I10; I20; J13

Keywords: Language skills, fertility, health, education, natural experiment

* *Corresponding author. IZA and Department of Economics, University of Aberdeen, Dunbar Street, Aberdeen AB24 3QY, UK.*
Acknowledgments: The permission of the Office for National Statistics (ONS) to use the Longitudinal Study is gratefully acknowledged, as well as the help and support of Nicky Rogers and the Longitudinal Study Development Team at ONS. The authors alone are responsible for the interpretation of the data.

1 Introduction

The foreign-born share of the population increased in almost all OECD countries between the decade from 2000 to 2010 (OECD, 2012), and the social integration of immigrants is a current priority on the policy agenda of developed countries. In order to implement successful policies to target social and health inequalities among their immigrant population, policy makers need to understand what barriers immigrants face to integrate. Among possible barriers, one is language. Language facilitates access and use of public services, such as those related to health and education, and this in turn may affect health and the educational achievement of immigrants. There is extensive evidence that better language skills improve immigrants' economic status, in particular their earnings, but there is limited research on how language affects their social life and family structures (Chiswick and Miller, 2014). There is also limited knowledge on how language affects immigrants' health outcomes and behaviour. This paper aims to contribute to this knowledge by identifying the causal effect of English language skills on a number of fertility, health and education outcomes for immigrants in England and Wales.

Our paper contributes to the literature on the effect of language skills on these social outcomes in a number of ways. First, we use a unique dataset from the Office for National Statistics England and Wales Longitudinal Study (LS) that links individual-level dataset from the 2011 Census for England and Wales and Live Births to Sample Mothers (LBSM) that contains information on births to LS sample women. The combination of these two datasets allows us to study the impact of language skills on various fertility outcomes that, to the best of our knowledge, have not been studied before, namely, a woman's age at having her first child, the number of children she has, and the birthweight of her children. Second, we are first to provide evidence on how language skills affect health outcomes in England and Wales. The study of the relationship between language skills and health outcomes in the United Kingdom (UK) is very limited because there is almost no health dataset collected in the UK that also incorporates information on language proficiency (Jayaweera, 2014). Third, by analysing data for England and Wales, we provide an important contribution to the literature by presenting results from countries with a different immigration composition to that of the United States (US), which is the country that has been most extensively studied. OECD (2012) indicates that the UK and US have similar shares of immigrants —11.3% of the total population in the UK, 12.5% in the US— but they are different in a key characteristic of interest to our analysis: 47% of immigrants in the UK come from a country with English as an official language, compared to 20% in the US. In addition, 47% of immigrants in the UK are highly educated, compared to 34% in the US, and 34% of immigrants in the UK come from an OECD high-income country, compared to only 14% in the US.

Credibly identifying and quantifying the impact of language proficiency on fertility, health, and education outcomes poses a significant empirical challenge because English language proficiency is endogenous. First, unobserved heterogeneity across individuals that affects both English proficiency and these social outcomes, such as ability and cultural attitude, may bias estimates of the effect of English proficiency. Second, these social outcomes can also affect an individual's English proficiency (reverse causality); for example, having children might improve English skills if the mother starts interacting more with other parents, schoolteachers, and healthcare professionals, but it could also have the opposite effect, if the mother quits her job or starts staying home for longer hours. Third, measurement errors in the measure of English proficiency can also cause a bias in the Ordinary Least Squares (OLS) estimator.

To address the endogeneity problem, we use an instrumental variable (IV) strategy where age at arrival in the UK is exploited to construct an instrument for English skills. Bleakley and Chin (2004) is the first paper that exploits age at arrival to construct an IV for language skills of immigrants based on the "critical period hypothesis" of language acquisition proposed by Lenneberg (1967). The hypothesis states that a person exposed to a language within the critical period of language acquisition (i.e., childhood) learns the language more easily, implying that immigrants who arrive in the UK at a younger age have on average better English language skills than those who arrive when they are older.

However, age at arrival alone is not a valid instrument because it is likely to have direct effects on the social outcomes of immigrants through channels different from language acquisition; for example , through cultural assimilation or better knowledge of UK institutions and social services, such as education and healthcare systems. To address this concern, we use immigrants from English-speaking countries as a control to partial out age-at-arrival effects that could affect the social outcomes of immigrants through channels different from language acquisition. More precisely, conditional on individual characteristics, any difference in the outcomes of early and late arrivers from English-speaking countries would reflect age-at-arrival effects, while the corresponding difference for immigrants from non-English-speaking countries would reflect both age-at-arrival effects and language effects. Thus, a difference in the outcomes between early and late arriver immigrants from non-English-speaking countries in excess of the corresponding difference for those from English-speaking countries can arguably be attributed to the effects of language. Based on this idea, we construct an IV which is an *interaction* of age at arrival and an indicator for coming from non-English-speaking countries.

The results obtained in our IV estimations indicate that better English-language skills considerably delay the age at which women have their first child, lower their likelihood of becoming a teenage mother, and decrease the number of children a woman gives birth to, but do not affect self-reported health and child health measured by child's birthweight. The impact of better En-

glish skills on educational achievement is also considerable: better English skills significantly raise the probability of having academic degrees and significantly lower the probability of having no qualifications.

The remainder of the paper proceeds as follows. Section 2 reviews the literature on the effect of language skills on social outcomes of immigrants. Section 3 presents our econometric specification and discusses empirical problems and our identification strategy. Section 4 describes our sample and data on fertility, health, and education, while main empirical findings are discussed in Section 5. Section 6 investigates the robustness of our main results to different sample and regression specifications. Finally, Section 7 discusses policy implications and concludes the paper.

2 Literature Review

The literature that explores the causal effect of language skills on health and fertility outcomes is not extensive. The role of language skills has been analysed by social scientists across numerous disciplines, including sociologists, epidemiologists and behavioural scientists, and their studies typically examine the correlation between language skills and health or fertility outcomes.

A small number of studies investigate the relationship between language skills and fertility. Focusing on individuals in the US with Hispanic origin, Lichter et al. (2012), Gorwaney et al. (1991) and Swicegood et al. (1988) examine the relationship between English proficiency and fertility. They conclude that poor English language proficiency is significantly associated with higher fertility rates among individuals with Hispanic origin. In contrast, evidence from Canada provides a different picture: Adsera and Ferrer (2014) analyse the relationship between language proficiency, age at arrival, and fertility patterns among Canadian natives and immigrants, using language fluency measured by whether the mother tongue of the immigrant is one of the Canadian official languages, English or French. Their results suggest that the number of children of immigrants increases with age at immigration relative to that of natives regardless of language proficiency. In other words, fertility of immigrants with English or French as their mother tongue is also higher than that of natives, implying that language proficiency is unlikely to play a key role in explaining a higher fertility among immigrants.[1] A possible issue with these studies is that unobserved heterogeneity that affects the fertility decision of a woman, such as cultural attitude, may be correlated with her language proficiency. Reverse causality may also

[1]It is worth noting that, in the study of Adsera and Ferrer (2014), the reference point is fertility of native-born Canadians, while in earlier studies using US data the comparison is made between immigrants with different degrees of language skills.

be an issue. Bleakley and Chin (2010) address this potential endogeneity using an interaction between age at arrival and coming from non-English-speaking countries as an IV for language skills of immigrants in the US. Their results suggest that the mother's English proficiency significantly reduces the number of children living in her household. A limitation of this study is that the number of children living in a household is not necessarily the actual number of children a woman has had.

Regarding health outcomes, numerous studies analyse the role of language skills in the context of acculturation in the US (Bauer et al., 2012; Kimbro et al., 2012; Lee et al., 2013; Miranda et al., 2011). Their findings appear to be mixed. Kimbro et al. (2012) and Miranda et al. (2011) find a positive association between English language proficiency and health outcomes, while Bauer et al. (2012) and Lee et al. (2013) find that this correlation is insignificant. Among the very few studies based on countries other than the US, Ng et al. (2008) and Ng et al. (2011) investigate the effect of proficiency in the official languages in Canada, English and French, on self-reported health. Their findings indicate that limited language proficiency in an official language is positively associated with poor self-reported health. An issue with these studies is that it is not clear if poor language skills deteriorate health due to, for example, a poor interaction with healthcare professionals, or if poor health hinders the development of language skills due to, for example, a limited interaction with people. Guven and Islam (2015) address this endogeneity issue using an interaction between age at arrival in Australia and coming from non-English-speaking countries as an IV for language skills. Their results indicate that better English skills improve self-reported health and physical health, but have an insignificant effect on mental health.[2] Clarke and Isphording (2015) address the issue of endogeneity using age at arrival in Australia and the linguistic background of the immigrant to construct an IV for language skills, and find that English language deficiency significantly deteriorates the physical health of immigrants.

Turning to educational outcomes, the relationship between language acquisition and education has been explored in few studies, that analyse the factors that explain the academic performance of immigrants. For example, Glick and White (2003) find that having a non-English background is associated with lower test scores of immigrants in the US. The bulk of studies that explore the educational attainment of immigrants do not focus directly on language proficiency but on age at arrival of immigrants and how it affects their ability to catch up with natives and second generation immigrants (e.g., Böhlmark 2008; Cortes 2006; Heckman 2001; Ohinata and van Ours 2012). Some of these studies suggest that language proficiency might be a key factor explaining their results; for example, Corak (2011) finds a negative impact of age at arrival on

[2]When the sample is divided by sex, the effect on self-reported health becomes negative and insignificant for men and remains positive and significant for women.

holding a high school diploma for immigrants who arrived in Canada after age nine, but only for those arriving from non-English or non-French speaking countries. Cohen Goldner and Epstein (2014) use data from Israel and arrive to a similar conclusion: age at arrival has a negative impact on the probability of graduating from high school. They suggest that a possible channel could be language acquisition. A challenge for studying the effect of language skills on education is that causation is difficult to establish because of the endogeneity of language skills. For instance, better language skills help achieve better academic results, but a higher level of education would also help improve language ability through, for example, a more frequent exposure to reading or writing in English. To overcome the endogeneity of language skills, Bleakley and Chin (2004) and Akbulut-Yuksel et al. (2011) use an interaction between age at arrival and coming from non-English-speaking countries as an IV for language skills. Both studies find that better English skills raise the number of years of schooling of immigrants in the US.

3 Identification Strategy

We explore the causal effect of English language proficiency on fertility, health and education outcomes of immigrants living in England and Wales by regressing these outcomes on a measure of English language proficiency, controlling for various individual characteristics. The following model is specified:

$$outcome_{ica} = \beta_0 + \beta_1 proficiency_{ica} + X'_{ica}\delta + \gamma_c + \eta_a + \epsilon_{ica} \tag{1}$$

where $outcome_{ica}$ represents the outcome of individual i born in country c who arrived in the UK at age a, and $proficiency_{ica}$ is a measure of English language proficiency. The individual characteristics, X_{ica}, and the parameter δ are $K \times 1$ vectors, where K is the number of variables capturing individual characteristics such as age and gender. γ_c and η_a are country-of-birth and age-at-arrival fixed effects, respectively, and ϵ_{ica} is the disturbance term.

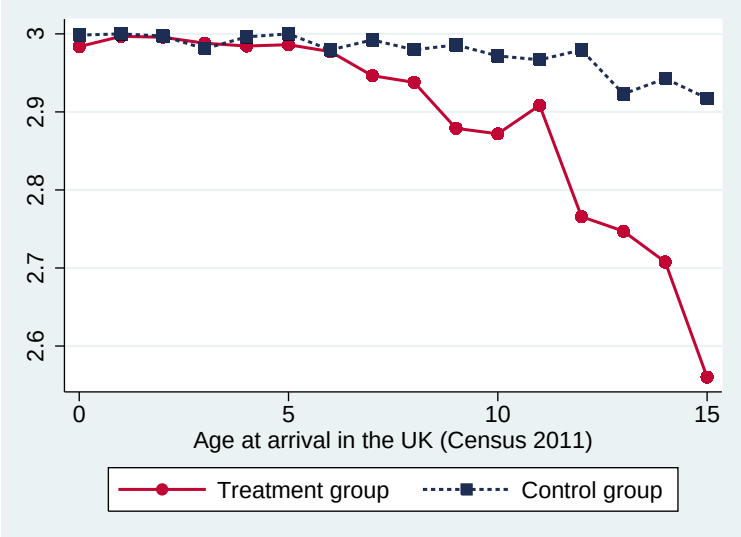
The main coefficient of interest is β_1, which measures the effect of English language proficiency on the outcomes analysed. An econometric issue in the estimation of equation (1) is the endogeneity of English language proficiency. First, unobserved heterogeneity, such as ability and cultural attitude, is likely to be correlated with both English language skills and the outcomes we study. For example, an individual with high ability is likely to attain higher levels of education, but at the same time may be more capable of learning to speak and write English proficiently. It is also plausible that a high ability individual has better health in part because he has better access to information on the consequences of his behaviour on smoking and drinking. If this is the case, language proficiency will be positively correlated with educational attainment

and health even if language proficiency does not "cause" an increase in educational attainment and an improvement in health. Second, fertility, health and education outcomes of an individual may affect language proficiency of the individual (reverse causality). For example, if one has bad health, she may not improve her language skills because her health problems may limit her interactions with other people. It could also be the case that having children improves language skills because it increases contacts with native speakers such as schoolteachers and healthcare professionals. Thus, it is hard to conclude if the social outcomes affect language proficiency or vice versa. Third, the measure of language proficiency, which is self-reported in our analysis, may contain measurement errors. For example, Dustmann and van Soest (2001), based on German data, find that a self-reported measure of language proficiency contains a substantial amount of measurement errors. For these reasons, the OLS estimator for β_1 is likely to be biased and inconsistent.

To identify the causal effect of language skills, we use an IV strategy, which requires an IV giving exogenous variation in English language skills. In this paper, we exploit age at arrival in the UK to construct an IV for language skills. The idea of using age at arrival in a host country to construct an IV for language proficiency is proposed by Bleakley and Chin (2004). Their idea is based on the hypothesis suggested by cognitive scientists, referred to as the critical period hypothesis: namely, if individuals are exposed to a new language at a critical age range (i.e., childhood), they can learn the language easily and at the level of natives, while acquiring a new language is much harder if individuals are first exposed to it after this critical age range (i.e., adults and adolescents). The critical period hypothesis implies that age at arrival in the UK would affect English language proficiency of immigrants arriving from countries where English is not spoken as a main language because these immigrants are exposed to English for the first time when they arrive in the UK. More specifically, for immigrants arriving from non-English-speaking countries, those who arrive at an early age are likely to easily learn English, while late arrivers would face more difficulties in acquiring English and may have a poorer command of English. In contrast, for immigrants arriving from English-speaking countries, age at arrival would not affect their proficiency in English because they had already been speaking English prior to their arrival in the UK.

For a variable to be a valid IV for English language skills, we require the assumption that the variable does not appear in equation (1) and is not correlated with any other determinants of immigrant's social outcomes than language skills. However, age at arrival is unlikely to satisfy this assumption for various reasons. First, age at arrival would affect not only language proficiency but also cultural assimilation in other aspects than language. For example, fertility rates of women in some countries such as Kenya, which account for a significant proportion of immigrants in the UK, are on average higher than those of UK-born women. Immigrants who

Figure 1: Age at arrival and English proficiency



Notes: Figure plots the average ordinal measure of English proficiency, where 3, 2, 1, and 0 correspond to speaks "very well", "well", "not well", and "not at all", respectively. The solid and dotted lines correspond to immigrants from English- and non-English-speaking countries, respectively.
Source: Authors' calculations based on the dataset from the Office for National Statistics England & Wales Longitudinal Study.

arrive in the UK at an early age from these higher-fertility countries might have low fertility rates because early arrivers are affected by UK cultural norms. Second, age at arrival would also increase knowledge about UK institutions, which may subsequently affect social outcomes of immigrants. For example, early arrivers may have an advantage over late arrivers in attaining a higher level of education because they are familiar with the UK educational systems. Likewise, early arrivers might have better health partly because they have a better knowledge of the UK healthcare systems.

To address these concerns, instead of using age at arrival as an IV, we use an *interaction* of age at arrival with a dummy variable for coming from a non-English-speaking country. All immigrants are exposed to a new environment at arrival in the UK irrespective of their country of origin, but only those coming from non-English-speaking countries encounter a new language. Thus, conditional on individual characteristics, differences in outcomes of early and late arrivers from English-speaking countries would reflect age-at-arrival effects only, whereas differences in outcomes of those from non-English-speaking countries would reflect both language effects and age-at-arrival effects. Therefore, a difference in the outcomes between early- and late-arriver immigrants from non-English-speaking countries in excess of the corresponding difference for those from English-speaking countries can be arguably attributed to the effects of language.

Figure 1 shows the relationship between age at arrival and English language proficiency among childhood immigrants in England and Wales. The solid and dotted lines correspond to immigrants from English- and non-English-speaking countries, respectively. The graph shows that, irrespective of age at arrival, immigrants from English-speaking countries are generally proficient in English (i.e., scoring between 2.9 and 3 in the ordinal measure of English proficiency, where 3 corresponds to "speaks very well"). This is not surprising because those from English-speaking countries were exposed to English prior to their arrival in the UK. In contrast, immigrants coming from non-English-speaking who arrived after age eight report having a poorer command of English. The two series start diverging at around age nine and for those arriving from non-English-speaking countries, the later they arrive, the poorer their English is. This is consistent with the critical period hypothesis. The pattern observed in Figure 1 leads us to parametrise age at arrival of individual i born in country c who arrived in the UK at age a, θ_{ica}, in the following way:

$$\theta_{ica} = max(0, arrival_i - 8) \times I(i\, coming\, from\, a\, non-English-speaking\, country) \quad (2)$$

where $arrival_i$ is age at arrival for individual i and $I(\cdot)$ is an indicator function that equals one if the individual comes from a non-English-speaking country, and zero otherwise. $max(0, arrival_i - 8)$ measures the distance from age eight for those arrived in the UK after age eight, and zero otherwise. An assumption underlying equation (2) is that there is no difference in English language proficiency between immigrants from English- and non-English-speaking countries for those who arrived at age eight or before, but language proficiency and age at arrival are linearly related after age eight for immigrants coming from non-English-speaking countries.[3] Using equation (2), the relationship between English language proficiency and age at arrival, which corresponds to our first-stage equation, can be specified as follows:

$$proficiency_{ica} = \alpha_0 + \alpha_1 \theta_{ica} + X'_{ica}\zeta + \iota_c + \kappa_a + u_{ica} \quad (3)$$

where the individual characteristics, X_{ica}, and the parameter ζ are $K \times 1$ vectors, where K is the number of variables capturing individual characteristics. ι_c and κ_a are country-of-birth and age-at-arrival fixed effects, respectively, and u_{ica} is the disturbance term.

For this IV strategy to identify the causal effect of language skills, we require the assumption that immigrants from English- and non-English-speaking countries are exposed to the same age-at-arrival effects except for language. However, one could question the credibility of this as-

[3] Our results are not sensitive to this particular cut-off at age eight.

sumption. Namely, the two sets of immigrants might face different age-at-arrival effects because they have different background characteristics. For example, a significant proportion of immigrants from non-English-speaking countries come from European countries such as Germany. These European countries have close economic and political ties and cultural commonalities with the UK due to, for example, the existence of the European Union and a long history of economic, political and cultural interactions, potentially making it easier for them to adapt to the new UK environment. Likewise, a significant proportion of immigrants from English-speaking countries come from Commonwealth countries. These countries also share some commonalities with the UK regarding, for example, culture and legal systems, possibly making it easier for them to adapt to the UK living environment. As long as these country-of-origin specific effects do not vary across age at arrival, these effects will be absorbed by country-of-origin fixed effects in equation (1). Nevertheless, one may still be concerned that these country-of-origin specific effects might vary across age at arrival. These concerns are further discussed in Section 6, where we conduct a series of robustness checks to address these issues.

4 Data and Sample

Data

We use data from the Office for National Statistics England and Wales LS, an individual-level dataset comprising linked census and life event records for 1% of the population of England and Wales. We make use of two datasets that are part of the LS: the 2011 Census for England and Wales and the LBSM, which contains information of live births in England and Wales to women usually resident in England and Wales between 1971 to 2011, taken from the birth registration and birth certificate.[4] We use data on live births to sample mothers that is contained in the LS to create our fertility outcomes for the mothers in our sample. As measures of fertility outcomes, the following variables are exploited: birthweight of child, age of mother when the first child was born, a dummy for whether the mother was a teenager when her first child was born, and the number of children born to a woman. The last variable is a better measure of the actual number of children born to a woman than the usual census variable of number of dependent children living in the same household and used in other studies that analyse census data such as Bleakley and Chin (2010).

Our measures on education and health are also obtained from the 2011 Census. We construct

[4]The dataset contains a variable that records the number of children previously born alive to sample mother. Prior to May 2012, this information was only collected for births within marriage. The registrar records the number of the mother's previous live born children by her present husband and any former husband. Therefore some births may not be recorded or only recorded if the mother gave this information to the registrar.

our set of education indicator variables from one single question in the 2011 Census, which collects self-reported information on the highest level of education achieved by an individual. The 2011 Census also collects information on self-reported health, which is an ordinal measure ranging between 1 (very bad health) to 5 (very good health). From this variable, we have derived the indicator variables "good or very good health" and "bad or very bad health". In addition, as another measure of health, an indicator variable for self-reported long term health problems is exploited.

The information on language skills and individual characteristics is also obtained from the 2011 Census. Using information on self-reported language skills, we construct our measure of English language skills, where 3, 2, 1, and 0 correspond to speaks English "very well", "well", "not well", "not at all". To create our instrument for language skills, information on the country of birth and age at arrival of immigrants are used.[5] The data on origin-country characteristics used in the section of robustness checks are from the following sources: the education datasets used are from Barro and Lee (2013), and all other country characteristics are from the World Development Indicators 2015[6].

Sample

Our empirical analysis is based on the sample of individuals in the LS dataset who (i) lived in England and Wales at the 2011 Census, (ii) are childhood immigrants, and (iii) were aged 25 to 60 at the 2011 Census. The minimum age restriction of 25 is imposed to allow individuals enough time to complete their education, while the maximum age restriction of age 60 is imposed to deal with the issue of selective mortality. Childhood immigrant is defined as an individual born outside of the UK who arrived in the UK for the first time at age 15 or before. At this age, we assume that immigrants did not make their own migration decisions but followed their parents or guardians who migrated to the UK. In our analysis of fertility outcomes, we further restrict this sample to females that have at least one child registered in the LBSM dataset.

In order to implement our identification strategy, we divide the sample into three mutually exclusive groups: (i) individuals born in countries where English is not an official language, (ii) individuals born in countries where English is an official language and the predominant language spoken, and (iii) individuals born in countries where English is an official language but not the predominant language spoken.[7] The first group is our "treatment group" and the second group

[5]The age of arrival in the UK is derived from the date that a person last arrived to live in the UK and their age. Short visits away from the UK are not counted in determining the date that a person last arrived. The age of arrival is only applicable to usual residents who were not born in the UK and does not include usual residents born in the UK who have emigrated and since returned.

[6]Downloaded from: `http://data.worldbank.org/data-catalog/world-development-indicators`.

[7]To categorise countries, we have used the World Almanac and Book of Facts 2011.

is our "control group". The third group is excluded from our sample because it is not clear to what extent individuals in this group were exposed to English prior to their arrival in the UK. Following this rule, some of immigrants who account for significant proportions of immigrants in the UK, such as Indians and Pakistanis, are excluded from the sample. The list of country of birth incorporated in our sample by treatment status can be found in Table 10.

Tables 1 and 2 present summary statistics separately for early and late arrivers in the UK. An individual is classified as an early arriver if he arrived in the UK at age eight or earlier. The cut-off value of eight is chosen because the average English proficiency of immigrants arriving from English- and non-English-speaking countries starts diverging at age at arrival nine (see Figure 1). This implies that, conditional on individual characteristics, those who arrived in the UK at age eight or earlier speak English as fluent as those from English-speaking countries when adults irrespective of age at arrival. After this age at arrival, English language proficiency decreases almost monotonically. In fact, English language proficiency, presented in Table 1, panel A, indicates that there is no significant difference between early arrivers coming from English- and non-English-speaking countries, while late arrivers coming from non-English-speaking countries appear to have a poorer command of English than those from English-speaking countries when adults. Another noticeable feature shown in Table 1 is that the shares of individuals coming from European and Commonwealth countries are different between the treatment and control groups. Of particular note, a majority of immigrants from English-speaking countries come from Commonwealth countries, while a majority of immigrants from non-English-speaking countries come from Europe. This is notably different from the case of US immigrants studied by, for example Bleakley and Chin (2010), where a significant proportion of immigrants from non-English-speaking countries come from Mexico.

5 Results

We begin by estimating equation (1) using the OLS estimator. Table 3 reports the OLS estimates of the effect of English language proficiency on social outcomes of childhood immigrants in England and Wales, after controlling for individual characteristics and country-of-birth and age-at-arrival fixed effects. Panels A to C of Table 3 present the results for fertility, health and education outcomes, respectively. The sample in panel A is restricted to mothers.

Panel A shows that better English proficiency is significantly associated with delayed fertility, a lower likelihood of becoming a teenage mother, and having fewer children (rows A1 to A3). Specifically, a one-unit increase in our English language ordinal measure (e.g., shifting from speaks English "not well" to "well") is significantly associated with a delay in the age of giving birth to the first child of approximately 2.6 years, a 0.13 lower probability of becoming

Table 1: Immigrant characteristics

	Arrived aged 0-8		Arrived aged 9-15	
	Born in English-speaking country	Born in non-English-speaking country	Born in English-speaking country	Born in non-English-speaking country
A. Individual characteristics				
English proficiency	2.99	2.98	2.95	2.75
ordinal measure	(0.09)	(0.16)	(0.23)	(0.53)
Age	44.03	40.83	46.58	37.80
	(9.70)	(10.53)	(10.93)	(10.60)
Female	0.51	0.51	0.54	0.49
	(0.50)	(0.50)	(0.49)	(0.50)
White	0.65	0.79	0.27	0.49
	(0.48)	(0.41)	(0.45)	(0.50)
Black	0.14	0.03	0.35	0.14
	(0.35)	(0.17)	(0.48)	(0.35)
Asian	0.16	0.08	0.34	0.19
	(0.37)	(0.27)	(0.47)	(0.40)
Other single race	0.01	0.06	0.01	0.13
	(0.09)	(0.24)	(0.09)	(0.33)
Multiracial	0.03	0.03	0.03	0.03
	(0.17)	(0.18)	(0.16)	(0.17)
Commonwealth	0.68	0.05	0.82	0.05
	(0.47)	(0.22)	(0.38)	(0.22)
Europe	0.20	0.65	0.09	0.33
	(0.40)	(0.48)	(0.30)	(0.47)

Notes: Standard deviations are shown in parenthesis. The sample consists of individuals in the ONS LS dataset that were present in the 2011 Census for England and Wales, are childhood immigrants, and were aged 25 to 60 at Census 2011. We define childhood immigrant as those individuals born outside of the UK that arrived in the UK for the first time at age 15 or earlier. Column (1) provides statistics for individuals, in the pre-treatment category, born in countries where English is an official language and the predominant language spoken (control group), while column (2) provides statistics for individuals, in the pre-treatment category, born in countries where English is not an official language (treatment group). Columns (3) and (4) provide statistics for the same groups but for the post-treatment category. An individual is classified into the pre-treatment category if he arrived in the UK at age eight or earlier. The observation numbers for panel A in columns (1) and (2) correspond to 2,932 and 2,188, respectively; the observation numbers in columns (3) and (4) correspond to 1,865 and 1,260 individuals, respectively.
Source: Authors' calculations based on the dataset from the Office for National Statistics England and Wales Longitudinal Study.

Table 2: Immigrant outcomes

	Arrived aged 0-8		Arrived aged 9-15	
	Born in English-speaking country	Born in non-English-speaking country	Born in English-speaking country	Born in non-English-speaking country
B. Education				
No qualifications	0.08	0.09	0.12	0.20
	(0.28)	(0.28)	(0.32)	(0.40)
Compulsory level qualification	0.38	0.39	0.43	0.44
	(0.49)	(0.49)	(0.49)	(0.49)
Post-compulsory level qualification	0.62	0.60	0.57	0.55
	(0.49)	(0.49)	(0.49)	(0.49)
Academic degree	0.45	0.42	0.38	0.36
	(0.50)	(0.49)	(0.48)	(0.48)
C. Health				
Self-reported health ordinal measure	4.23	4.26	4.11	4.23
	(0.87)	(0.88)	(0.88)	(0.86)
Good or very good health	0.85	0.85	0.80	0.84
	(0.36)	(0.36)	(0.40)	(0.37)
Bad or very bad health	0.04	0.05	0.05	0.04
	(0.21)	(0.22)	(0.22)	(0.20)
Long-term health problem	0.12	0.13	0.15	0.12
	(0.33)	(0.33)	(0.36)	(0.32)
D. Fertility (women aged 25 and over)				
Age at having first child	27.28	26.39	26.50	24.85
	(5.34)	(5.11)	(5.33)	(5.51)
Teenage mother	0.11	0.12	0.11	0.16
	(0.31)	(0.33)	(0.32)	(0.37)
Number of children born to mother	2.24	2.22	2.40	2.40
	(0.91)	(0.94)	(1.17)	(1.08)
Birthweight of child (grammes)	3293.93	3365.99	3171.0	3317.9
	(588.75)	(571.73)	(563.1)	(533.3)

Notes: Standard deviations are shown in parenthesis. The sample consists of individuals in the ONS LS dataset that were present in the 2011 Census for England and Wales, are childhood immigrants, and were aged 25 to 60 at Census 2011. We define childhood immigrant as those individuals born outside of the UK that arrived in the UK for the first time at age 15 or earlier. Column (1) provides statistics for individuals, in the pre-treatment category, born in countries where English is an official language and the predominant language spoken (control group), while column (2) provides statistics for individuals, in the pre-treatment category, born in countries where English is not an official language (treatment group). An individual is classified into the pre-treatment category if he arrived in the UK at age eight or earlier. The observation numbers for panels B and C in columns (1) and (2) correspond to 2,932 and 2,188, respectively; the observation numbers in columns (3) and (4) correspond to 1,865 and 1,260 individuals, respectively. The sample for fertility outcomes (Panel D) consists of childhood immigrant females aged 25 and over; sample sizes in Panel D vary by outcome: birthweight (1,851; 1,311; 1,103, and 633 in columns (1) to (4), respectively), age at which the woman had her first child (636; 433; 332; 186), teenage mother (1,005; 731; 647; 339), number of children (710; 491; 421; 238).

Source: Authors' calculations based on the dataset from the Office for National Statistics England and Wales Longitudinal Study.

Table 3: OLS estimates of the effects of English proficiency

Dependent variable		English proficiency	Standard errors
A. Fertility	A1. Age at having first child	2.647***	(0.517)
	A2. Teenage mother	-0.130***	(0.034)
	A3. Number of children	-0.437***	(0.135)
	A4. Birth weight	18.85	(28.65)
B. Health	B1. Self-reported health	0.365***	(0.044)
	B2. Good health	0.151***	(0.017)
	B3. Bad health	-0.048***	(0.016)
	B4. Long-term health problem	-0.123***	(0.021)
C. Education	C1. No qualifications	-0.282***	(0.021)
	C2. Compulsory-level qualification	-0.229***	(0.021)
	C3. Post-compulsory qualification	0.235***	(0.020)
	C4. Academic degree	0.232***	(0.016)

Notes: *** $p<$ .01. Standard errors are clustered by country of birth. Controls included in the analysis are dummy variables for sex, Commonwealth origin, European origin, race, age, age at arrival, and country of origin. The full sample is used for the analyses in panels B and C, where the sample size, N, is 8,245. The sample is restricted to mothers in row A2 (N = 2,722). The sample is further restricted to mothers whose information about the first child is available in row A1 (N = 1,588) and to mothers whose complete number of children is known in row A3 (N = 1,861). Row A4 uses dataset at child level where N = 4,898 (i.e., the mother appears multiple times in the dataset in case she gave birth multiple times).

Source: Authors' calculations based on the dataset from the Office for National Statistics England and Wales Longitudinal Study.

a teenage mother, and giving birth to 0.44 less children on average. However, English skills appear to have no significant association with child health measured by birthweight (row A4). Turning to health outcomes for adults, panel B indicates that better English proficiency is significantly correlated with better self-reported health (rows B1 and B2) and a lower likelihood of reporting bad or very bad health and having long-term health problems (rows B3 and B4). For example, row B2 indicates that a one-unit increase in English skills significantly increases the probability of reporting very good or good health by approximately 0.15 on average. Regarding educational outcomes, panel C shows that better language skills are positively correlated with the likelihood of obtaining a higher level of educational qualifications. Specifically, better language skills are significantly associated with a lower probability of having no qualifications or having only compulsory-level qualifications (rows C1 and C2), and are significantly associated with a higher probability of having a post-compulsory qualification and an academic degree (rows C3 and C4). For example, a one-unit increase in our English language ordinal measure is significantly correlated with an increase in the probability of having an academic degree by approximately 0.23 on average (row C4).

However, the OLS estimator of β_1 in equation (1) is biased if (i) unobserved heterogeneity across individuals that affects our social outcomes, such as ability and cultural attitude, is also correlated with fluency in English, (ii) the social outcomes and English skills are simultaneously determined, and/or (iii) the English proficiency measure is correlated with measurement errors. To address this potential endogeneity of English skills, we estimate equation (1) using the IV estimator, where we use the *interaction* of age-at-arrival and a dummy variable for coming from non-English-speaking countries as an instrument for English skills.[8] Table 4 presents the first-stage and reduced-form estimates of the effects of the instrument on English skills and on our social outcomes, respectively, and the IV estimates of the effects of English skills on the social outcomes (i.e., β_1 in equation (1)). Panels A to C correspond to the regressions for fertility, health and education outcomes, respectively. The sample is restricted to mothers in panel A.

The first-stage estimates presented in panels B and C of column (1) indicate that, for those from non-English-speaking countries, each year past age eight at arrival significantly decreases our English language skills ordinal measure by approximately 0.04 on average. When the sample is restricted to mothers in panel A, the coefficient estimates increase in absolute terms and range between -0.06 and -0.07. It might be the case that English proficiency of females is more sensitive to age at arrival. The magnitude of the coefficient implies that a person's English ordinary measure would be approximately lower by half a unit if the person arrives from non-English-speaking countries at age 15 instead of at age eight.

[8] Precisely, the instrument equals the excess age at arrival from age eight for those who arrived from non-English-speaking countries, and zero otherwise.

Table 4: First-stage, reduced-form, and IV estimates

Dependent variable:	English proficiency	Fertility, health or education	
	First-stage	Reduced-form	IV
	(1)	(2)	(3)
A. Fertility			
A1. Age at having first child	-0.055***	-0.214**	3.864**
	(0.020)	(0.093)	(1.882)
A2. Teenage mother	-0.064***	0.013**	-0.210**
	(0.016)	(0.007)	(0.094)
A3. Number of children	-0.063***	0.045*	-0.718*
	(0.015)	(0.023)	(0.373)
A4. Birthweight of child	-0.069***	0.637	-9.216
	(0.018)	(6.890)	(100.2)
B. Health			
B1. Self-reported health	-0.040***	-0.009	0.221
	(0.011)	(0.008)	(0.177)
B2. Good health	-0.040***	-0.002	0.041
	(0.011)	(0.003)	(0.074)
B3. Bad health	-0.040***	-0.000	0.006
	(0.011)	(0.002)	(0.041)
B4. Long-term health problem	-0.040***	-0.003	0.071
	(0.011)	(0.003)	(0.073)
C. Education			
C1. No qualifications	-0.040***	0.022***	-0.537***
	(0.011)	(0.005)	(0.072)
C2. Compulsory	-0.040***	0.007	-0.184
	(0.011)	(0.006)	(0.122)
C3. Post-compulsory	-0.040***	-0.008	0.192
	(0.011)	(0.006)	(0.120)
C4. Academic degree	-0.040***	-0.015*	0.372***
	(0.011)	(0.008)	(0.138)

Notes: *** $p<.01$, ** $p<.05$, and * $p<.10$. Standard errors are clustered by country of birth. First-stage and reduced-form estimates are the estimated coefficients on the dummy variable for late arrivers (i.e., those arriving after age eight) coming from non-English-speaking countries. The IV estimates are the estimates of α_1 in equation (1). Rows in each panel correspond to the regressions for the different measures of fertility, health and education in panels A, B, and C, respectively. Refer to Table 3 for the controls included and sample sizes.
Source: Authors' calculations based on the dataset from the Office for National Statistics England and Wales Longitudinal Study.

Regarding fertility outcomes reported in panel A, the reduced-form estimates presented in column (2) show that, for each year at arrival past age eight, the age at which the mother has her first child significantly decreases (row A1), and both the probability of becoming a teenage mother and the number of children a mother gives birth to significantly increase (rows A2 and A3). The causal effects of interest presented in column (3) show that a one-unit increase in English skills significantly raises the mother's age at which she has her first child by approximately 3.9 years (row A1), and significantly lower her likelihood of becoming a mother in her teens by approximately 0.21 (row A2). In addition to the timing of having a child, English proficiency also affects the number of children a woman gives birth to: a one-unit increase in our English skill measure significantly reduces the number of children a woman has by approximately 0.72 (row A3). This is a sizable effect corresponding to a reduction of approximately 68 per cent relative to the mean value for childhood immigrants who arrived after age eight from non-English-speaking countries. We did not find any effect of English skills on child health, measured by birthweight.

Turning to health outcomes for adults reported in panel B, the reduced-form estimates show that arriving after age eight has no significant effect on any of the self-reported health measures we analyse. In line with the reduced-form estimates, the IV estimates presented in column (3) show that better English skills have no significant effect on self-reported health. Compared to the corresponding OLS estimates in Table 3 that show significant associations between English skills and self-reported health, the magnitudes of IV estimates are lower in absolute terms. A possible interpretation is that unobserved individual heterogeneity that is correlated with both English language proficiency and self-reported health, such as ability, biases the OLS estimator upwards. For example, an individual with a higher ability may be capable of learning a new language quickly and may also have a good health condition due to, for instance, better knowledge about the consequences of their behaviour or better earning potentials in the labour market. If this is the case, language proficiency can be positively correlated with health even if language proficiency does not "cause" an improvement in health.

Panel C reports educational outcomes. The reduced-form estimates in column (2) show that, after age eight, each additional year that passes before an individual arrives in the UK increases his likelihood of having no qualifications or having only compulsory-level qualifications (rows C1 and C2), and decreases his likelihood of obtaining post-compulsory qualifications and academic degrees (rows C3 and C4), although the estimates for compulsory-level and post-compulsory qualifications are insignificant. The causal effects of interest reported in column (3) indicate that better English language skills significantly lower the probability of having no qualifications and raise that of obtaining academic degrees (rows C1 and C4). The IV estimates are larger than the corresponding OLS estimates in absolute terms, almost double the size of the

OLS estimate for the probability of having no qualifications. The point estimates suggest that a one-unit increase in English language skills lowers the probability of having no qualifications by 0.54 and raises the probability of obtaining academic degrees by 0.37, both sizable effects. Because understanding the language used at school is likely to be a key component of academic success, it is not surprising that individuals with better English skills have a lower probability of having no qualifications and a better probability of obtaining academic degrees. Regarding the likelihood of obtaining only compulsory-level qualifications or post-compulsory-level qualifications, the IV estimates in rows C2 and C3 are insignificant. Taken together, our findings suggest that proficiency in English affects the likelihood of having the highest and the lowest levels of educational attainment (i.e., no qualifications and academic degrees), but has no effect on the likelihood on the educational attainment at a medium level.

Mechanisms at work

We have found that better English proficiency significantly affects fertility outcomes and the educational attainment of immigrants. Having estimated the effects of English proficiency, in this subsection we explore the possibility that education mediates the effects of language proficiency on fertility outcomes. We do this by controlling for measures of education, in addition to English proficiency, in our fertility regressions. It might be the case that better English skills improve educational attainments and career opportunities for females, which in turn may delay the timing in which a woman has her first child or reduce the number of children she has. As measures of education, we include dummy variables that equal one if the person has no qualifications, a post-compulsory qualification, or an academic degree, respectively, and zero otherwise. The reference group is individuals with compulsory education. A caveat is that estimates of the effects of English proficiency on fertility outcomes no longer have causal interpretations because education is likely to be endogenous. Despite this limitation, we present results as suggestive evidence of the possible role that education plays in determining fertility outcomes.

Even-numbered columns of Table 5 present the effects of English proficiency on the age at which a woman has her first child, her probability of becoming a mother in her teens, the number of children she has, and her children's birthweight, respectively, after controlling for education. The base results from Table 4 without controlling for education are copied to adjacent odd-numbered columns for comparison purposes. Column (2) indicates that the point estimate of the effect of English skills on age at which the mother had her first child is greatly reduced by nearly 55 per cent relative to the corresponding estimate in column (1), and is no longer statistically significant. Likewise, column (4) shows that the point estimate for the likelihood of becoming a teenage mother is lowered by 25 per cent compared to the corresponding estimate in column (3),

Table 5: The effects of education and language on fertility

Dependent variable:	Age at having first child		Teenage mother		Number of children		Birthweight	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
English skills	3.864**	1.733	-0.210**	-0.157	-0.718*	-0.551	-9.216	-5.785
	(1.882)	(1.819)	(0.094)	(0.099)	(0.373)	(0.409)	(100.2)	(114.5)
No qualifications		-2.072***		0.073**		0.328**		-6.039
		(0.718)		(0.032)		(0.135)		(53.37)
Post-compulsory		0.512		-0.051***		-0.025		-9.203
		(0.416)		(0.019)		(0.061)		(30.81)
Academic degree		2.247***		-0.039***		-0.054		9.256***
		(0.358)		(0.014)		(0.066)		(26.91)
Education controls	no	yes	no	yes	no	yes	no	yes
# Observations	1,588	1,588	2,722	2,722	1,861	1,861	4,898	4,898

Notes: *** $p<$ .01 and ** $p<$ .05. Standard errors are clustered by country of birth. The estimates are the IV estimates of α_1 in equation (1). Refer to Table 3 for the controls included in the analyses except for even-numbered columns, where dummy variables for having no qualifications, a post-compulsory qualification, and an academic degree are additionally controlled for.
Source: Authors' calculations based on the dataset from the Office for National Statistics England and Wales Longitudinal Study.

and becomes insignificant. In contrast, education significantly affects the age at which a woman has her first child and her likelihood of becoming a teenage mother. For example, relative to individuals with compulsory-level qualifications, women with no qualifications and academic degrees have their first child approximately 2.1 years earlier and 2.2 years later, respectively. Similar results hold for the regression on the number of children a woman has: after controlling for education, the point estimate is lowered by nearly 25 per cent, and becomes insignificant (column (6)). The results provide some evidence in favour of the argument that a key channel through which English proficiency affects the fertility decisions of immigrant women is education: an improvement in language skills results in a higher educational attainment, which in turn could be delaying the age at which women have their first child and the number of children they have. The results are different from the findings based on US evidence where the effects of language proficiency remain significant even after controlling for measures of education (see Bleakley and Chin, 2010), implying that mechanisms through which English language skills affect fertility outcomes might be different between the UK and the US.

Turning to child health measured by child birthweight, the effect of English proficiency after controlling for education remains insignificant (column (8)). An interesting point to note is that having degrees appears to affect child health measured by birthweight: relative to those with compulsory-level qualifications, children of mothers with academic degrees weight approximately 93 grammes more on average.

6 Robustness Checks

In this section, we address the concern that immigrants from English- and non-English-speaking countries could be different in aspects other language that could also affect their social outcomes; if this was the case, then immigrants from English-speaking countries would not be a good control in our estimation. To address this concern, we employ two different strategies: (i) we consider different sample specifications in which we restrict our sample to immigrants from countries that are less likely to be heterogeneous, and (ii) we control for an interaction of age at arrival with different origin-country characteristics.

A key assumption for our IV strategy to credibly identify the causal effects of language skills is that immigrants from English- and non-English-speaking countries are exposed to the same age-at-arrival effects aside from language. Under this assumption, immigrants from English-speaking countries can be used to control for age-at-arrival effects that immigrants from non-English-speaking countries are also exposed to. However, one may cast doubt on the validity of this assumption: English-speaking countries might be economically, culturally, and institutionally more similar to the UK, making this age-at-arrival effects different for immigrants from these two groups of countries. More precisely, even after excluding language differences, immigrants from non-English-speaking countries might face a larger barrier to adapt to the UK environment. As long as these country-of-origin specific effects do not vary across age at arrival, these effects will be absorbed by country-of-origin fixed effects in equation (1).

However, these country-of-origin specific effects could vary across age at arrival. For example, for those who arrive in the UK at an early age, country-of-origin specific characteristics may not affect their social outcomes in the UK because they left their origin country sufficiently early not to be affected by their origin-country characteristics. In contrast, those who arrive in the UK at a later age might be affected by their origin-country characteristics because they are exposed to those characteristics for a longer period of time. This type of concerns may be less severe in the UK context than in other contexts, for example the US context where most studies of similar nature are based. For example, 47% of UK immigrants are highly educated, compared to 34% in the US, and 34% of UK immigrants come from an OECD high-income country, compared to only 14% of immigrants in the US (OECD, 2012). Thus, the average characteristics of immigrants from the two groups of countries might be more similar in the case of UK immigrants than in the case of US immigrants.

Nevertheless, to address this type of concerns, we retain in our sample only immigrants from countries that might be less heterogeneous from each other. In particular, we exclude from our sample immigrants from Europe and Commonwealth countries in columns (2) and (3) of Table 6, respectively. The base results from the previous section are copied to column (1). As shown in

Table 1, significant proportions of immigrants from English- and non-English-speaking countries come from Commonwealth and European countries, respectively. The European countries have close economic and political ties and cultural similarities with the UK due to, for example, the existence of the European Union and a long history of economic, political and cultural interactions. Likewise, Commonwealth countries also share some commonalities with the UK regarding, for example, culture and legal systems. Omitting immigrants from these countries that might have special ties with the UK may make the retained countries more similar to each other.

Estimation results for fertility outcomes summarised in panel A are robust to changes in sample specifications: proficiency in English delays the age at which a woman has had first child, lowers her likelihood of becoming a teenage mother, and reduces the number of children she has, but has an insignificant effect on child health. A difference to be noted is that standard errors increase when European countries are omitted from the samples (column (1)), and the estimates become insignificant. This is not surprising as the sample sizes are reduced by omitting European countries. However, the point estimates reported in column (2) are not significantly different from our base results. Estimation results for health outcomes summarised in panel B show that the effects of English skills on health outcomes remain insignificant even after restricting our samples.

Turning to educational outcomes reported in panel C, the results are qualitatively similar to our main findings, although several interesting differences arise. After restricting the sample, the effects of English language proficiency increase in magnitude (in absolute terms) for all outcomes. Furthermore, the effects on the probability of having compulsory-level and post-compulsory level qualifications become significant, implying that English language proficiency has greater effects for immigrants coming from countries that might be less similar to the UK than European and Commonwealth countries. A possible interpretation is that the educational systems in Europe and Commonwealth countries might be more similar to those in the UK, making it easier for immigrants from these countries to adapt to the UK educational systems, irrespective of their proficiency in English language.

We now address in a different way this concern that immigrants from English- and non-English-speaking countries are not exposed to the same non-language age-at-arrival effects, by controlling for interactions of age at arrival with various origin country characteristics. Unless otherwise stated, we use origin country characteristics in 1970.[9] We begin by discussing our results for fertility outcomes summarised in Table 7. Base results from the previous section are

[9]The year 1970 is chosen because the average age of the immigrants in our sample is 41 as of 2011, implying that the average immigrants were born around 1970. We also consider using the values in 1980 (i.e., a decade after the time of birth of the average immigrants in our sample). Our results are not sensitive to the choice of year.

Table 6: IV estimates using alternative sample specifications

	All (1)	No Europe (2)	No Commonwealth (3)
A. Fertility			
A1. Age at having first child	3.864**	2.738	3.956**
	(1.882)	(2.018)	(1.786)
A2. Teenage mother	-0.210**	-0.225	-0.201**
	(0.094)	(0.143)	(0.0866)
A3. Number of children	-0.718*	-0.632	-0.758**
	(0.373)	(0.423)	(0.337)
A4. Birthweight of child	-9.216	-23.54	109.1
	(100.2)	(130.2)	(106.2)
B. Health			
B1. Self-reported health	0.221	0.036	0.198
	(0.177)	(0.197)	(0.202)
B2. Good health	0.041	-0.035	0.070
	(0.074)	(0.074)	(0.083)
B3. Bad health	0.006	0.020	0.004
	(0.041)	(0.043)	(0.056)
B4. Long-term health problem	0.071	0.068	0.050
	(0.073)	(0.072)	(0.082)
C. Education			
C1. No qualifications	-0.537***	-0.487***	-0.571***
	(0.072)	(0.074)	(0.076)
C2. Compulsory	-0.184	-0.264**	-0.281***
	(0.122)	(0.131)	(0.106)
C3. Post-compulsory	0.192	0.274**	0.295***
	(0.120)	(0.126)	(0.102)
C4. Academic degree	0.372***	0.428***	0.516***
	(0.138)	(0.136)	(0.135)

Notes: *** $p<.01$, ** $p<.05$, and * $p<.10$. Standard errors are clustered by country of birth. The estimates shown are the IV estimates of α_1 in equation (1), using the controls specified in Table 3. The results shown in columns (1) to (3) correspond to different sample specifications: Full sample (column (1)), full sample excluding Europe (column (2)), and full sample excluding Commonwealth countries (column (3)). The number of observations that corresponds to each of these samples varies by outcome. For all education and health outcomes, the full sample contains 8,245 observations, the sample excluding Europe, 5,644, and the sample excluding Commonwealth countries, 5,526. For the fertility outcomes, the number of observations in each of these samples, respectively, is presented in parenthesis: Age at having first child (1,588; 1,034; 881); Teenage mother (2,722; 1,731;, 1530); Number of children (1,861; 1172, 1063); Birthweight of child (4,898; 3,236; 2666).
Source: Authors' calculations based on the dataset from the Office for National Statistics England and Wales Longitudinal Study.

copied to column (1). Column (2) controls for an interaction of age at arrival with total fertility rate in the country of origin. If immigrants from non-English-speaking countries, on average, come from lower (or higher) fertility countries than those from English-speaking countries *and* the effects of origin-country fertility rate vary by age at arrival in the UK, the interaction of age at arrival with coming from non-English-speaking countries (i.e., our instrument for English language proficiency) captures the compound effects of language proficiency and differential fertility rates in the origin country. Column (2) of Table 7 suggests that our results are not sensitive to the inclusion of the interaction of age at arrival with total fertility rate in origin country, except for the number of children a mother has. For this outcome, standard errors of the coefficient estimate on English proficiency increase and the estimate becomes insignificant. However, the point estimate is not significantly different from our base result reported in column (1).[10]

In a similar spirit, to account for infant health in origin countries that might differently affect non-language age-at-arrival effects faced by immigrants from our two groups of countries, in the regression for birthweight, we control for an interaction with infant mortality rate, a measure of infant health, in the origin country. Column (3) indicates that the estimate of the effect of English proficiency remains insignificant, confirming that English skills are unlikely to affect child health.

Turning to health outcomes, columns (2) to (3) in Table 8 control for interactions with per capita health expenditure and life expectancy in the country of origin, respectively. Base results from the previous section are copied to column (1). Note that the figures in 1995 are used for per capita health expenditure (the earliest year for which data is available). Table 8 shows that our results are not sensitive to the inclusion of these additional controls, confirming our previous findings that proficiency in English does not have significant effects on self-reported health outcomes.

Regarding educational outcomes, columns (2) and (3) of Table 9 control for interactions with average years of schooling and pupil-teacher ratio in secondary education in the country of origin, respectively. Base results from the previous section are copied to column (1). Columns (2) and (3) indicate that our results are not sensitive to the inclusion of the additional variables, confirming that better English skills significantly reduce the likelihood of having no qualifications. A difference to be noted is that standard errors of the effect of English skills on the likelihood of having academic degrees increase, and the effect is now imprecisely estimated (row 4). However, the point estimates are not significantly different from the corresponding estimate in column (1).

[10]In the regressions for mother's age at which she had her first child and her likelihood of becoming a teenage mother, we also estimated our results controlling for the interaction of age at arrival with adolescent fertility rate in the origin country, defined as the number of births per 1,000 women aged between 15 - 19. Our results are not sensitive to the inclusion of this additional control.

Table 7: IV estimates for fertility outcomes using alternative controls for origin country characteristics

	Control for country of origin characteristics		
	Base results	Fertility rate x age at arrival	Infant mortality rate x age at arrival
	(1)	(2)	(3)
Age at having first child	3.864**	4.024*	
	(1.882)	(2.418)	
Teenage mother	-0.210**	-0.218**	
	(0.094)	(0.096)	
Number of children	-0.718*	-0.526	
	(0.373)	(0.421)	
Birthweight of child	-9.216	-1.705	-5.376
	(100.2)	(121.2)	(174.3)

Notes: *** $p<$ .01, ** $p<$ .05, and * $p<$ .10. Standard errors are clustered by country of birth. The estimates shown are the IV estimates of α_1 in equation (1) for the outcomes indicated in each row, using the controls specified in Table 3 and the additional control for origin country characteristics specified in each column. Column 1 presents the base results. Columns 2 to 4 present results including an additional control variable each, that is the interaction of age at arrival with an origin country characteristic, in 1970: Total fertility rate (Column 2), GDP per capita (Column 3), and infant mortality rate (Column 4). This latter control only applies to the outcome Birthweight of child. The number of observations that corresponds to each of these outcomes and specifications varies by outcome and specification: Age at having first child (1,588; 1,309 for each specification shown in columns 1 to 2, respectively); Teenage mother (2,722; 2,239); Number of children (1,861; 1,536); Birthweight of child (4,898; 4,002; 4,272, for each specification shown in columns 1 to 3).

Source: Authors' calculations based on the dataset from the Office for National Statistics England and Wales Longitudinal Study.

Table 8: IV estimates for health outcomes using alternative controls for origin country characteristics

	Control for country of origin characteristics		
	Base results	Health expenditure x age at arrival	Life expectancy x age at arrival
	(1)	(2)	(3)
Self-reported health	0.221	0.710	0.222
	(0.177)	(0.509)	(0.207)
Good health	0.041	0.054	0.011
	(0.074)	(0.199)	(0.086)
Bad health	0.006	0.126	0.027
	(0.041)	(0.133)	(0.050)
Long-term health problem	0.071	0.148	0.082
	(0.073)	(0.222)	(0.084)

Notes: *** $p<.01$, ** $p<.05$, and * $p<.10$. Standard errors are clustered by country of birth. The estimates shown are the IV estimates of α_1 in equation (1) for the outcomes indicated in each row, using the controls specified in Table 3 and the additional control for origin country characteristics specified in each column. Column 1 presents the base results. Columns 2 to 3 present results including an additional control variable each that is the interaction of age at arrival with an origin country characteristic: Per capita health expenditure in 1995 (column 2) and life expectancy in 1970 (column 3). The number of observations that corresponds to each of the specifications in columns 1 to 3 varies by specification: Column 1 (8,245 observations), column 2 (6,494), and column 3 (7,666).

Source: Authors' calculations based on the dataset from the Office for National Statistics England and Wales Longitudinal Study.

Table 9: IV estimates for education outcomes using alternative controls for origin country characteristics

	Control for country of origin characteristics		
	Base results	Years of education x age at arrival	Pupil-teacher ratio x age at arrival
	(1)	(2)	(3)
No qualifications	-0.537***	-0.474***	-0.382***
	(0.072)	(0.113)	(0.119)
Compulsory	-0.184	-0.126	0.162
	(0.122)	(0.206)	(0.283)
Post-compulsory	0.192	0.148	-0.139
	(0.120)	(0.204)	(0.282)
Academic degree	0.372***	0.294	0.0985
	(0.138)	(0.241)	(0.305)

Notes: *** $p<$ .01, ** $p<$ .05, and * $p<$ .10. Standard errors are clustered by country of birth. The estimates shown are the IV estimates of α_1 in equation (1) for the outcomes indicated in each row, using the controls specified in Table 3 and the additional control for origin country characteristics specified in each column. Column 1 presents the base results. Columns 2 to 3 present results including an additional control variable each, that is the interaction of age at arrival with an origin country characteristic, in 1970: Average number of years of education of individuals aged 25 and over (column 2) and pupil-teacher ratio in Secondary school (column 3). The number of observations that corresponds to each of these specifications in columns 1 to 3 varies by specification: Column 1 (8,245 observations), column 2 (6,494), and column 3 (7,239).

Source: Authors' calculations based on the dataset from the Office for National Statistics England and Wales Longitudinal Study.

7 Conclusions

Policy at present stresses that English proficiency is key to the integration of immigrants in the UK, but there is little research evidence on how English skills affect fertility, health and educational attainment of the immigrant population in the UK. From an international perspective, the UK is a particularly interesting country for studying the phenomenon of assimilation because the immigrants' composition in the UK is very different from that of the US, the country that has been most extensively studied: 47% of the immigrants in the UK come from a country with English as an official language, compared to 20% in the case of the US; 47% of immigrants in the UK are highly educated, compared to 34% in the US, and 34% of immigrants in the UK come from an OECD high-income country, compared to only 14% of immigrants in the US.

In this paper, we study how English language skills affect fertility, health and education outcomes of childhood immigrants in England and Wales, using a unique dataset that links individual-level data from the 2011 Census for England and Wales and the LBSM. We study the causal effect of language skills using an IV estimation strategy where age at arrival in the UK is exploited to construct an instrument for language skills. The idea of using age at arrival to construct the instrument is based on the critical period of language acquisition hypothesis, stating that a person exposed to a language within the critical period of language acquisition (i.e., childhood) learns the language more easily. The hypothesis implies that, for immigrants arriving from countries where English is not spoken as a main language, those who arrive in the UK at a younger age have on average better English language skills than those who arrive when they are older.

We incorporate immigrants from English speaking countries in our analysis as a control to partial out age-at-arrival effects that may affect the social outcomes of immigrants through channels different from language acquisition. More precisely, conditional on individual characteristics, any difference in the outcomes of early and late arrivers from English-speaking countries would reflect age-at-arrival effects, while the corresponding difference for immigrants from non-English-speaking countries would reflect both age-at-arrival effects and language effects. Thus, a difference in the outcomes between early and late arrivers of immigrants from non-English-speaking countries in excess of the corresponding difference for those from English-speaking countries could arguably be attributed to the effects of language. Based on this idea, we construct an instrumental variable for English language skills by interacting age at arrival with an indicator variable for coming from non-English-speaking countries.

Our results, based on IV estimations, suggest that fertility and educational outcomes of immigrants are influenced by their ability to speak English. We find that better English language skills significantly delay the age at which immigrant women have their first child, lower their

likelihood of becoming mothers in their teens, decrease the number of children they give birth to, but have no effect on child health measured by their children's birthweight. We also find that immigrants who speak English proficiently are more likely to have academic degrees, as compared to immigrants not proficient in English, who are more likely to have no qualifications. Supplementary regressions to explore the potential role that education plays in determining fertility outcomes of immigrants suggest that a higher educational attainment, that is the result of improved English language skills, is likely to delay the timing of having child and the number of children a woman has.

This finding is different from the case of the US where language skills significantly affect the number of children living in same household even after controlling for education (see Bleakley and Chin, 2010), implying that the mechanism through which language skills affect fertility choice of women is likely to be different in the UK and the US. Regarding health outcomes, we did not find any significant effects of English skills on adult self-reported health.

Our results have important policy implications. First, giving support to immigrants to learn and improve their English language skills may allow them to better participate in the educational systems in England and Wales. This in turn may improve their educational attainment and affect their fertility choices. Second, specific English-learning programs at school for young immigrants that arrived in the UK at age eight or later would help them improve their English language skills, since having arrived in the country after the critical period of language acquisition makes it more difficult for them to learn the language. A language learning support could have an important impact in their capability to obtain an education qualification and further pursue higher degree studies.

Although the primary focus of this paper is immigrants, the relevance of the paper is not limited to them. A better integration of the immigrant population would make them more productive, which in turn benefits the society as a whole. For example, the UK health care sector is greatly supported by workers from overseas as is shown by the fact that approximately 14 per cent of professionally qualified clinical staff are foreign nationals (Health and Social Care Information Centre, 2014). The British Medical Association states that many services provided by the National Health Service would struggle to provide effective care for their patients without a contribution of non-British staff. A better integration of immigrants may thus benefit society by securing an inflow of workers in this sector as well as by improving the quality of care provided by migrant carers.

Table 10: Individuals by country of birth

English-speaking countries			Non-English-speaking countries		
Country	Observations	Percent	Country	Observations	Percent
Kenya	747	15.5	Germany	966	28.0
Ireland	623	12.9	Cyprus	316	9.2
Jamaica	503	10.4	Turkey	178	5.2
Uganda	321	6.7	Somalia	151	4.4
South Africa	275	5.7	Italy	134	3.9
Singapore	244	5.1	Vietnam	132	3.8
United States	236	4.9	Malaysia	104	3.0
Canada	211	4.4	Portugal	81	2.3
Australia	207	4.3	Iran	80	2.3
Nigeria	194	4.0	Yemen	72	2.1
Malta	150	3.1	France	71	2.1
Tanzania	146	3.0	Malawi	60	1.7
Zambia	109	2.3	Iraq	56	1.6
Zimbabwe	108	2.2	Netherlands	46	1.3
Ghana	106	2.2	China	46	1.3
Guyana	79	1.6	Spain	45	1.3
New Zealand	69	1.4	Afghanistan	44	1.3
Gibraltar	50	1.0	Egypt	42	1.2
Trinidad and Tobago	47	1.0	Libya	35	1.0
Mauritius	39	0.8	Belgium	35	1.0
Sierra Leone	34	0.7	Poland	33	1.0
St Kitts and Nevis	31	0.6	Saudi Arabia	32	0.9
Barbados	28	0.6	Morocco	29	0.8
Isle Of Man	27	0.6	Lebanon	29	0.8
St Vincent and the Grenadines	26	0.5	Kosovo	29	0.8
St Lucia	24	0.5	Switzerland	27	0.8
Grenada	23	0.5	Thailand	26	0.8
Montserrat	21	0.4	Bahrain	25	0.7
Jersey	19	0.4	Sweden	24	0.7
Dominica	14	0.3	Ethiopia	21	0.6
Total Top 20	4,711	97.9	Total Top 20	2,969	86.0

Notes: Number of individuals by country of birth for the top 20 countries present in our sample that are English-speaking countries (control group, columns 1 to 3) and that are non-English speaking countries (treatment group, columns 4 to 6).

Source: Authors' calculations based on the dataset from the Office for National Statistics England and Wales Longitudinal Study.

References

Alicia Adsera and Ana Ferrer. Factors influencing the fertility choices of child immigrants in canada. *Population Studies*, 68(1):65–79, 2014.

Mevlude Akbulut-Yuksel, Hoyt Bleakley, and Aimee Chin. The effects of english proficiency among childhood immigrants: Are hispanics different? In David L. Leal and Stephen J. Trejo, editors, *Latinos and the Economy*, Immigrants and Minorities, Politics and Policy, pages 255–283. Springer New York, 2011.

Robert Barro and Jong-Wha Lee. A new data set of educational attainment in the world 1950–2010. *Journal of Development Economics*, 104(C):184–198, 2013.

Amy M. Bauer, Chih-Nan Chen, and Margarita Alegria. Prevalence of physical symptoms and their association with raceethnicity and acculturation in the united states. *General Hospital Psychiatry*, 34(4):323 – 331, 2012.

Hoyt Bleakley and Aimee Chin. Language skills and earnings: Evidence from childhood immigrants. *The Review of Economics and Statistics*, 86(2):481–496, 2004.

Hoyt Bleakley and Aimee Chin. Age at arrival, english proficiency, and social assimilation among us immigrants. *American Economic Journal: Applied Economics*, 2(1):165–92, 2010.

Anders Böhlmark. Age at immigration and school performance: A siblings analysis using swedish register data. *Labour Economics*, 15(6):1366 – 1387, 2008.

Barry R. Chiswick and Paul W. Miller. International Migration and the Economics of Language. In Barry R. Chiswick and Paul W. Miller, editors, *Handbook on the Economics of International Migration*. Elsevier, 2014.

Andrew Clarke and Ingo E. Isphording. Language Barriers and Immigrant Health Production. IZA discussion paper 8846, Institute for the Study of Labour, 2015.

Sarit Cohen Goldner and Gil S Epstein. Age at immigration and high school dropouts. *IZA Journal of Migration*, 3(1), 2014.

Miles Corak. Age at immigration and the education outcomes of children. IZA Discussion Papers 6072, Institute for the Study of Labor (IZA), 2011.

Kalena E. Cortes. The effects of age at arrival and enclave schools on the academic performance of immigrant children. *Economics of Education Review*, 25(2):121–132, 2006.

Christian Dustmann and Arthur van Soest. Language Fluency And Earnings: Estimation With Misclassified Language Indicators. *The Review of Economics and Statistics*, 83(4):663–674, 2001.

JenniferE. Glick and MichaelJ. White. academic trajectories of immigrant youths: Analysis within and across cohorts. *Demography*, 40(4):759–783, 2003.

N. Gorwaney, M. D. Van Arsdol, D. M. Heer, and L. A. Schuerman. Migration from latin american countries to the united states: The economic, social and reproductive lives of hispanic female immigrants, 1980. *International Migration*, 29(4):573–599, 1991.

Cahit Guven and Asadul Islam. Age at Migration, Language Proficiency, and Socioeconomic Outcomes: Evidence From Australia. *Demography*, 52(2):513–542, 2015.

James J. Heckman. Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *Journal of Political Economy*, 109(4):pp. 673–748, 2001.

Sarah Janssen. *The World Almanac and Book of Facts 2011*. Infobase Learning, New York, 2010.

Hiranthi Jayaweera. Health of migrants in the uk: What do we know? The migration observatory, University of Oxford, 2014. Accessed on 11 December 2014.

Rachel Tolbert Kimbro, Bridget K. Gorman, and Ariela Schachter. Acculturation and self-rated health among latino and asian immigrants to the united states. *Social Problems*, 59(3):pp. 341–363, 2012.

S Lee, A ONeill, ES Ihara, and DH Chae. Change in self-reported health status among immigrants in the united states: Associations with measures of acculturation. *PLoS ONE*, 8(10): 76494, 2013.

Eric H. Lenneberg. *Biological foundations of language*. Wiley, New York, 1967.

Daniel T. Lichter, Kenneth M. Johnson, Richard N. Turner, and Allison Churilla. Hispanic assimilation and fertility in new u.s. destinations. *International Migration Review*, 46(4): 767–791, 2012.

Patricia Y. Miranda, Hector M. Gonzalez, and Wassim Tarraf. Pathways between acculturation and health: Does the measure matter? *Hispanic Journal of Behavioral Sciences*, 33(4):524–539, 2011.

Edward Ng, Kevin Pottie, Denise Spitzer, Alia Mohammed, and Richard Glazier. Language proficiency, gender and self-reported health: An analysis of the first two waves of the longitudinal survey of immigrants to canada. *Canadian Journal of Public Health*, 99(6):505–510, 2008.

Edward Ng, Kevin Pottie, and Denise Spitzer. Official language profi ciency and self-reported health among immigrants to canada. Health Reports 4, Statistics Canada, 2011.

OECD. Settling in: Oecd indicators of immigrant integration 2012. 2012. Accessed on 11 December 2014.

Asako Ohinata and Jan C. van Ours. Young immigrant children and their educational attainment. *Economics Letters*, 116(3):288 – 290, 2012.

Gray Swicegood, Frank D. Bean, Elizabeth Hervey Stephen, and Wolfgang Opitz. Language usage and fertility in the mexican-origin population of the united states. *Demography*, 25(1): pp. 17–33, 1988.

CINCH working paper series

1 Halla, Martin and Martina Zweimüller**. Parental Responses to Early Human Capital Shocks:** Evidence from the Chernobyl Accident. CINCH 2014.

2 Aparicio, Ainhoa and Libertad González. **Newborn Health and the Business Cycle:** Is it Good to be born in Bad Times? CINCH 2014.

3 Robinson, Joshua J. **Sound Body, Sound Mind?:** Asymmetric and Symmetric Fetal Growth Restriction and Human Capital Development. CINCH 2014.

4 Bhalotra, Sonia, Martin Karlsson and Therese Nilsson. **Life Expectancy and Mother-Baby Interventions:** Evidence from A Historical Trial. CINCH 2014.

5 Goebel, Jan, Christian Krekel, Tim Tiefenbach and Nicolas R. Ziebarth. **Natural Disaster, Environmental Concerns, Well-Being and Policy Action:** The Case of Fukushima. CINCH 2014.

6 Avdic, Daniel, **A matter of life and death? Hospital Distance and Quality of Care:** Evidence from Emergency Hospital Closures and Myocardial Infarctions. CINCH 2015.

7 Costa-Font, Joan, Martin Karlsson and Henning Øien. **Informal Care and the Great Recession.** CINCH 2015.

8 Titus J. Galama and Hans van Kippersluis. **A Theory of Education and Health.** CINCH 2015.

9 Dahmann, Sarah. **How Does Education Improve Cognitive Skills?:** Instructional Time versus Timing of Instruction. CINCH 2015.

10 Dahmann, Sarah and Silke Anger. **The Impact of Education on Personality:** Evidence from a German High School Reform. CINCH 2015.

11 Carbone, Jared C. and Snorre Kverndokk. **Individual Investments in Education and Health.** CINCH 2015.

12 Zilic, Ivan. **Effect of forced displacement on health.** CINCH 2015.

13 De la Mata, Dolores and Carlos Felipe Gaviria. **Losing Health Insurance When Young:** Impacts on Usage of Medical Services and Health. CINCH 2015.

14 Tequame, Miron and Nyasha Tirivayi. **Higher education and fertility:** Evidence from a natural experiment in Ethiopia. CINCH 2015.

15 Aoki, Yu and Lualhati Santiago. **Fertility, Health and Education of UK Immigrants:** The Role of English Language Skills. CINCH 2015.

DOI: 10.17185/duepublico/70919
URN: urn:nbn:de:hbz:464-20191206-154438-6