Fertility, Health and Education of UK Immigrants: The Role of English Language Skills[*†]

Yu Aoki[‡] and Lualhati Santiago[§]

April 2015

<Preliminary version. Please do not cite without authors' permission.>

Abstract

This paper aims to identify the causal effects of English language skills on fertility, health and education outcomes of immigrants in England and Wales. To identify the causal effects, 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 easily. This implies that immigrants who arrive in the UK at a young age will 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 women have their first child, lower the likelihood of becoming a teenage mother, decrease the number of children a woman has, but do not affect child's birthweight and self-reported health. The impact on educational achievement is also considerable: better English skills significantly raise the probability of obtaining post-compulsory qualifications and academic degrees and significantly lower the probability of having no qualifications or only compulsory level qualifications.

Keywords: Language skills, fertility, health, education, natural experiment; JEL: I10, I20, J13

[*]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.

[†]This work contains statistical data from the ONS which is Crown Copyright and all statistical results remain Crown Copyright. The use of the ONS Statistics statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

[‡]IZA and Department of Economics, University of Aberdeen, Dunbar Street, AB24 3QY, United Kingdom.

[§]Public Policy Division, Social and Analysis Directorate, Office for National Statistics, Segensworth Road, Titchfield, PO15 5RR, United Kingdom.

1. Introduction

The foreign-born share of the population has increased in almost all OECD countries between 2000/01 and 2009/10 (OECD, 2012), and the social integration of immigrants is high 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, this paper focuses on language. Language facilitates access and use of public services, such as 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 & Miller, 2014). There is also limited knowledge of 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: 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 relation 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 proficiency in English 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. 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 & Chin (2004) propose using 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). This 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 all age-at-arrival effects that would 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 reflects age-at-arrival effects, while any difference in the outcomes in the case of immigrants from non-English-speaking countries reflects 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 be arguably attributed to the effects of language. Based on this idea, we construct an instrumental variable 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 child's birthweight and self-reported health. The impact of better English skills on education achievement is also considerable: better English skills significantly raise the probability of hav-

ing post-compulsory qualifications and academic degrees and significantly lower the probability of having no qualifications or only compulsory-level 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 discusseed in Section 5. Section 6 investigates the robustness of our main results to different sample and regression specifications. Finally, Section 7 concludes the paper and discusses our plans for further investigation.

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 a 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. Their results indicate that poor English proficiency is significantly associated with higher fertility rates among individuals with Hispanic origin. In contrast, evidence from Canada provides a different picture: Adsera & 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 fertility 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 native born, implying that language proficiency is unlikely to play a key role in explaining a higher fertility among immigrants.[1] A caveat in these studies is the endogeneity of language skills; for example, unobserved heterogeneity that affects the fertility decision of a woman, such as cultural attitude, may also be correlated with her language proficiency. Reverse causality may also be an issue. Bleakley & 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 pro-

[1] It is worth noting that, in the study of Adsera & Ferrer (2011), the reference point is fertility of native born Canadians unlike earlier studies using US data where comparison is made among immigrants with different degrees of language skills.

ficiency 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 the actual number of children a woman has had. We overcome this limitation by using information on the actual number of children a woman has given birth to, contained in the ONS LS dataset.

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. There are 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 official language proficiency 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 & Islam (2013) address the endogeneity issue of language skills 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, but have an adverse effect on mental health and an insignificant effect on physical health.[2]

The relation between language acquisition and education of immigrant children has been explored in studies that analyse the factors that explain the academic performance of immigrants and their children. In the US, Glick & White (2003) find that having a non-English background is associated with lower test scores of immigrants, and Portes & MacLeod (1999) find that parental knowledge of English language is positively correlated with better academic performance of their children. The bulk of studies that explore the education attainment of immigrants do not focus directly on language proficiency but on age at arrival of immigrant children and how it affects their ability to catch up with native and second generation immigrants.[3] Some of these studies hypothesise that language proficiency might be a key factor explaining their results; for example, Corak (2011) exploits Canadian 2006 Census data and finds a negative impact of age at arrival on

[2]When the sample is divided by sex, the effect on physical health becomes significant at a 10 per cent level for a male sample, while the effect on mental health becomes insignificant for both female and male samples.

[3]For example, a positive relation between arriving at an early age and achieving more years of schooling is found based on US data (Chiswick & DebBurman, 2004; Gonzalez, 2003; Heckman, 2001; Perreira et al., 2006), Norwegian data (Bratsberg et al., 2011), Israeli data (Cohen Goldner & Epstein, 2014), and Canadian data (Corak, 2011; Schaafsma & Sweetman, 2001). A positive relation between arriving at an early age and test scores at school is found in Cortes (2006) for the US, Böhlmark (2009) and Böhlmark (2008) for Sweden, and Ohinata & van Ours (2012) for the Netherlands.

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 & 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 may be language acquisition. Based on data from Sweden, Böhlmark (2008) also finds a negative impact of age at arrival on school performance after age nine.

A problem of studying the relation between language skills and education is that causation is difficult to establish due to 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. To overcome the endogeneity of language skills, Bleakley & 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 of immigrants in the US. They find that a one-unit increase in English skills raises years of schooling by about two years (Bleakley & Chin, 2004) or three years (Akbulut-Yuksel et al., 2011).

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:

$$outcome_{ica} = \alpha_0 + \alpha_1 proficiency_{ica} + X'_{ica}\delta + \gamma_c + \eta_a + \varepsilon_{ica} \tag{1}$$

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 time varying 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. γ_c and η_a are country-of-birth and age-at-arrival fixed effects, respectively, and ε_{ica} is the disturbance term.

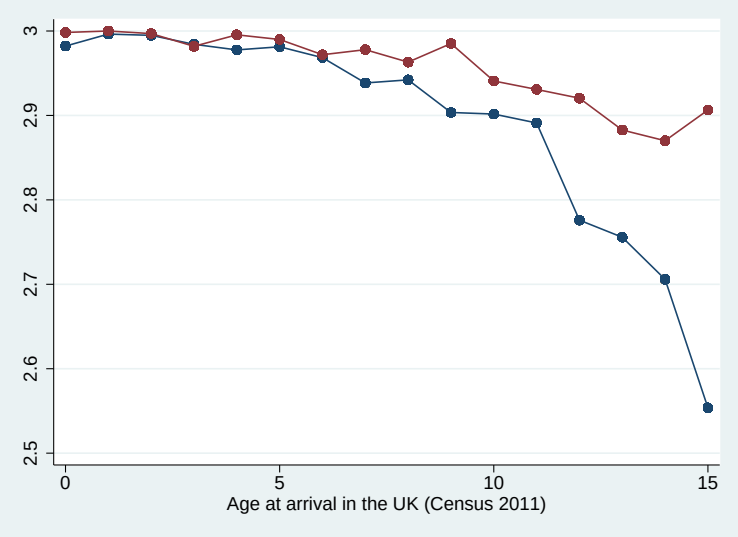
The main coefficient of interest is β_1, which measures the effect of English language proficiency on the outcomes analysed. An econometric issue in the estimation of equation (1) is the endogeneity of English language proficiency. First, unobserved heterogeneity, such as ability, is likely to be correlated with both English language skills and our outcomes. For example, an individual with a high ability is likely to attain a high level of education, and at the same time may acquire English more easily. It is also plausible that a high ability individual has a good health

condition due to, for instance, better knowledge about the consequences of their behaviour such as 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 be limiting her interactions with English speakers. It could also be the case that having children improves language skills because it increases her English-speaking contacts, for example schoolteachers and healthcare professionals. Thus, it is hard to conclude if health and fertility affect language proficiency or vice versa. For these reasons, the OLS estimator for β_1 is unlikely to estimate the causal effect of language proficiency on these outcomes.

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 & Chin (2004). Their idea is based on the hypothesis proposed by cognitive scientists, referred to as the critical period hypothesis: 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 proficiently, while late arrivers will face more difficulties for acquiring English and may have poorer English language skills. In contrast, for immigrants arriving from English-speaking countries, age at arrival would not affect their proficiency in English because they already spoke 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, conditional on language skills, the instrument has no direct effect on fertility, health and education outcomes. 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 India and Pakistan, which account for a significant proportion of immigrants in the UK, are on average higher than those of UK-born women. Immigrants who arrive in the UK at an early age from these higher-fertility countries might have low fertility rates because early arrivers are affected by cultural norms in the UK. Second, age at arrival would also increase knowledge about

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 red and blue 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.

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 evade these problems, instead of using age at arrival as an IV for English language skills, 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 outcome between immigrants from English- and non-English-speaking countries of the differences between early and late arrivers is 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 red and blue lines correspond to im-

migrants 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 the arrival in the UK. In contrast, among immigrants from non-English-speaking countries, those who arrived at or before age eight speak English as well as those arriving from English-speaking countries, while those 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 after age eight, the later they arrive, the poorer their English is. This is consistent with the critical period hypothesis. The pattern observed in Figure 1 motivates 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:

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

where *arrival* is age at arrival and $I(\cdot)$ is an indicator function that equals one if individual i comes from a non-English-speaking country, and zero otherwise. $max(0, arrival - 8)$ 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 or before age eight, but language proficiency and age at arrival are linearly related after age eight for immigrants coming from non-English-speaking countries. We will examine the sensitivity of our results to this assumption in section 6. 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:

$$proficiency_{ica} = \beta_0 + \beta_1 \theta_{ica} + X'_{ica}\zeta + \iota_c + \kappa_a + u_{ica} \quad (3)$$

where the time varying 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 effects of language skills, we require the assumption that those 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 assumption. For example, a significant proportion of immigrants from non-English speaking countries come from European countries, including Poland and 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. Likewise, Commonwealth countries also share some commonalities with the UK regarding, for example, culture and legal systems. Thus, immigrants coming from European and Commonwealth countries might face different age-at-arrival effects from those coming from outside these countries. We therefore control for Europe and Commonwealth dummies in our model. In addition, in section 6, we exclude from our sample individuals that arrive from European and Commonwealth countries, as they could have more commonalities with the UK.

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 for 1971 to 2011 taken from the birth registration and birth certificate.[4] The 2011 Census for England and Wales contains information on self-reported language skills from which 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". The measure of English proficiency is identical to those used in studies of similar nature to ours.[5] The 2011 Census for England and Wales also includes information on the country of birth and age at arrival of immigrants, which allows us to create our instrument for language skills[6].

We evaluate the impact of language skills on health, fertility and education. Our measures on education and health are obtained from the 2011 Census for England and Wales. We construct our set of education indicator variables from one single variable in the 2011 Census for England and Wales, which collects self-reported information on the highest level of education achieved by the individual. The 2011 Census for England and Wales also collects information on self-reported

[4]The dataset contains a variable that records the number of children previously born alive to sample mother. Prior to April 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 mother gives this information to the registrar.

[5]Bleakley & Chin (2010) create a similar variable from the 2010 US Census of Population and Housing and argue that it is a credible measure of language skills for studying the relation between language and socioeconomic outcomes.

[6]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.

health, which is an ordinal measure that takes values 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, the 2011 Census for England and Wales collects information on long term health problems, also self-reported. We use data on live births to sample mothers that is collected by the LS to create our fertility outcomes for the mothers in our sample: birth weight 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 number of children born to the mother. This latter variable is a better measure of children born to a mother than the usual census variable of number of dependent children in the household used in other studies that analyse census data, such as Bleakley & Chin (2010).

Sample

Our empirical analysis is based on the sample of individuals in the LS dataset who were living in England and Wales at the 2011 Census, are childhood immigrants and are currently aged 25 to 60. We define childhood immigrants as individuals born outside of the UK who arrived in the UK for the first time at age 15 or before. 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 our sample into three mutually exclusive groups: individuals born in countries where English is not an official language, individuals born in countries where English is an official language and the predominant language spoken, and individuals born in countries were English is an official language but not the predominant language spoken.[7] The first group is our "treatment group" and the second group is our "control group". We exclude the third group 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.

Table 1 presents summary statistics for our regressors and outcome variables for immigrants who arrived in the UK in the pre-treatment period. We classify an individual into the pre-treatment category 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 (c.f., Figure 1). This implies that, for those who arrived in the UK at age eight or earlier, conditional on individual characteristics, age at arrival has no effect on their English proficiency when adults. In panels B, C, and D, we present the outcome variables for immigrants that arrived in the UK before age nine. The data shows that there are no important differences between those born in English- and non-English-speaking countries. However, there are differences in some characteristics of these two groups, as can be seen in Panel A. In particular, racial composition and the share of individuals coming

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

Table 1: Immigrant characteristics

	(1)		(2)	
	Born in English-speaking country		Born in non-English -speaking country	
	mean	s.d.	mean	s.d.
A. Regressors				
English proficiency ordinal measure	2.99	0.09	2.98	0.16
Age	44.03	9.7	40.83	10.53
Female	0.51	0.5	0.51	0.5
White	0.65	0.48	0.79	0.41
Black	0.14	0.35	0.03	0.17
Asian / Pacific Islander	0.16	0.37	0.08	0.27
Other single race	0.01	0.09	0.06	0.24
Multiracial	0.03	0.17	0.03	0.18
Commonwealth	0.68	0.47	0.05	0.22
Europe	0.2	0.4	0.65	0.48
B. Education				
No qualifications	0.08	0.28	0.09	0.28
Compulsory-level qualification	0.38	0.49	0.39	0.49
Post-compulsory-level qualification	0.62	0.49	0.6	0.49
Academic degree	0.45	0.5	0.42	0.49
C. Health				
Self-reported health, ordinal measure	4.23	0.87	4.26	0.88
Good or very good health	0.85	0.36	0.85	0.36
Bad or very bad health	0.04	0.21	0.05	0.22
Long-term health problem	0.12	0.33	0.13	0.33
D. Fertility (women only)				
Age at having first child	27.28	5.34	26.39	5.11
Teenage mother	0.11	0.31	0.12	0.33
Number of children born to mother	2.24	0.91	2.22	0.94
Birthweight of child (grammes)	3293.93	588.75	3365.99	571.73

Notes: 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 are currently aged 25 to 60. 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 A to C in columns (1) and (2) correspond to 2,932 and 2,188, 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 in column(1), 1,311 in column(2)), age at which she had her first child (636, 433), teenage mother (1005, 731), number of children (710, 491).
Source: Authors' calculations based on the dataset from the Office for National Statistics England and Wales Longitudinal Study.

from European and Commonwealth countries are different in our two groups. We will explore these concerns in section 6, when we will conduct robustness checks incorporating as controls some potentially relevant country characteristics. The data presented in Section 6 comes from the following sources: The Penn World Tables version PWT8.1 (Barro & Lee, 2013; Feenstra et al., 2013; World Development Indicators 2015).[8]

5. Results

We begin by estimating equation (1) using the Ordinary Least Squares (OLS) estimator. Table 2 reports the OLS estimates of the effect of English language proficiency on social outcomes after controlling for individual characteristics and country-of-birth and age-at-arrival fixed effects, using data on childhood immigrants in England and Wales. Panels A, B and C of Table 2 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 having the first child of approximately 2.6 years, a 0.13 lower probability of becoming a teenage mother, and giving birth to 0.44 less children on average. However, English skills appear to have no significant association with a child's birthweight (row A4), which is a measure of child health. 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 education 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).

[8]In particular, the GDP per capita dataset used comes from,Feenstra et al. (2013), the education datasets used come from Barro & Lee (2013), and all other country characteristics come from the World Development Indicators 2015, downloaded from: `http://data.worldbank.org/data-catalog/world-development-indicators`.

Table 2: 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	0.364***	(0.044)
	B2. Good health	0.150***	(0.017)
	B3. Bad health	-0.048***	(0.016)
	B4. Long-term health problem	-0.122***	(0.021)
C. Education	C1. No qualifications	-0.282***	(0.021)
	C2. Compulsory-level qualification	-0.229***	(0.021)
	C3. Post-compulsory qualification	0.236***	(0.020)
	C4. Academic degree	0.231***	(0.017)

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 8,249. 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 C3 (N = 1,861). Row A4 uses dataset at a child level where N = 4,898 (i.e., the mother appears multiple times in the dataset in case she gave a birth multiple times).
Source: Authors' calculations based on the dataset from the Office for National Statistics England and Wales Longitudinal Study.

The OLS estimator 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, or (ii) our social outcomes and English skills are simultaneously determined. 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.[9] Table 3 presents the first-stage and reduced-form estimates of the effect of the instrument on English skills and on our social outcomes, respectively, and the IV estimates of the effect of English skills on our social outcomes (i.e., α_1 in equation (1)). Panels A, B, and C correspond to the regressions for fertility, health and education outcomes, respectively. The first-stage estimates presented in column (1) indicate that, for those from non-English speaking countries, each year past age eight at arrival significantly decreases our English language skill 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 females are more sensitive to age at arrival regarding English proficiency. 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.

Panel A reports fertility outcomes, the sample is restricted to mothers. 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 sheh as her first child by approximately 3.9 years (row A1), and significantly lower her likelihood of becoming a teenage mother by approximately 0.21 (row A2). In addition to the timing of having a child, English proficiency also affects number of children a woman gives birth to: a one-unit increase in our English skills' 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. It could be the case that better English skills improve the educational attainment and career opportunities for females, which in turn delays the time at which they have their first child and reduces the number of children they have. Regarding child birthweight, a measure for child health, there seems to be no significant effect of English skills on it.

[9] 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 3: 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. Birth weight	-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 2 for the controls included and the sample sizes.
Source: Authors' calculations based on the dataset from the Office for National Statistics England and Wales Longitudinal Study.

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 2 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 upward. For example, an individual with a high ability may acquire language quickly and have a good health condition at the same time 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.

Regarding educational outcomes reported in panel C, 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. Given that 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 chance of having no qualifications and a better chance of obtaining academic degrees. Regarding the likelihood of obtaining only compulsory-level qulifications or post-compulsory-level qualifications, the IV estimates in column (3) are insignificant (rows C2 and C3). Thus, 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.

Table 4: The effects of education and language on fertility

Dependent variable:	Age at having first child		Teenage mother		Number of children		Birth weight	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
English skills	3.858**	1.733	-0.209**	-0.157	-0.743**	-0.551	-9.216	-57.85
	(1.878)	(1.819)	(0.092)	(0.099)	(0.370)	(0.409)	(100.2)	(114.5)
No qualifications		-2.072***		0.073**		0.328**		-60.39
		(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		92.56***
		(0.358)		(0.014)		(0.066)		(26.91)
Education controls	no	yes	no	yes	no	yes	no	yes
# Observations	1,587	1,587	2,722	2,722	1,860	1,860	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 2 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.

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 delays the timing in which a woman has her first child or reduces 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 dummy variable for compulsory education is omitted from the regressions. Because education is likely to be endogenous, estimates of the effects of English proficiency on fertility outcomes no longer have causal interpretations. Despite this limitation, we present these results as suggestive evidence of the possible role that education plays in determining fertility outcomes.

Even-numbered columns of Table 4 present the effects of English proficiency on the age at which a woman has her first child, her probability of becoming a teenage mother, the number

of children she has, and child birthweight, respectively, after controlling for education. The main results from Table 3 without controlling for education are shown in 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), 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.

Turning to our child health measure, 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 first address the concern that, apart from language, immigrants from English- and non-English-speaking countries are different in aspects that might affect their social outcomes, which would imply that immigrants from English-speaking countries are not a good control in our model for immigrants from non-English speaking countries. 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 likely to be more similar to each other, and (ii) we control for different origin country characteristics.

A key assumption for our IV strategy to 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, apart from the language, immigrants from non-English-speaking countries might face a larger barrier to adapt to the new environment, and the severity of this barrier may increase as age at arrival increases. This concern might be less severe in our UK context than in other contexts, for example in the US context, on which most studies of similar nature are based. OECD (2012) indicates that 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. The average characteristics of immigrants from the two groups of countries might thus be more similar in the case of immigrants in the UK than in the case of immigrants in the US.

Nevertheless, to address this type of concern, we retain in our sample only immigrants from countries that might be more similar to each other. In particular, we exclude from our sample immigrants from Europe and Commonwealth countries in columns (2) and (3) of Table 5, 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, support our main findings in Table 3: proficiency in English delays the age at which a woman has hert 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 birthweight. 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, although the point estimates are not significantly different from those in our main findings. Estimation results for health outcomes, summarised in panel B, are also similar to our main findings in Table 3.

Turning to educational outcomes, reported in panel C, most of our results are similar to the main findings, although several interesting differences arise. After restricting the sample, the effects of English language proficiencyg 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 matters more 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

Table 5: IV estimates using alternative sample specifications

	All (1)	No Europe (2)	No Commonwealth (3)
A. Fertility			
A1. Age at having first child	3.689**	2.738	3.956**
	(1.780)	(2.018)	(1.786)
A2. Teenage mother	-0.194**	-0.225	-0.201**
	(0.0890)	(0.143)	(0.0866)
A3. Number of children	-0.704*	-0.632	-0.758**
	(0.356)	(0.423)	(0.337)
A4. Birthweight of child	-29.76	-23.54	109.1
	(99.89)	(130.2)	(106.2)
B. Health			
Self-reported health	0.231	0.0358	0.198
	(0.175)	(0.197)	(0.202)
Good health	0.0466	-0.0346	0.0701
	(0.0736)	(0.0739)	(0.0832)
Bad health	0.00688	0.0200	0.00393
	(0.0417)	(0.0434)	(0.0560)
Long-term health problem	0.0676	0.0683	0.0508
	(0.0720)	(0.0718)	(0.0821)
C. Education			
No qualifications	-0.521***	-0.487***	-0.571***
	(0.0689)	(0.0738)	(0.0764)
Compulsory	-0.167	-0.264**	-0.281***
	(0.123)	(0.131)	(0.106)
Post-compulsory	0.175	0.274**	0.295***
	(0.122)	(0.126)	(0.102)
Academic degree	0.368***	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 2.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 8,158 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 (1575, 1034, 881); Teenage mother (2701, 1731, 1530); Number of children (1844, 1172, 1063); Birthweight (4867, 3236, 2666).
Source: Authors' calculations based on the dataset from the Office for National Statistics England and Wales Longitudinal Study.

in Europe and Commonwealth countries may have a greater similarity with 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 adress the concern that immigrants from English- and non-English-speaking countries are not exposed to the same non-language age-at-arrival effects in a different way, by controlling for interactions of age at arrival with various origin country characteristics. Unless otherwise stated, we use origin country characteristics in 1970[10]. We begin by discussing our regressions for fertility outcomes, results are summarised in Table 6. Column (2) controls for an interaction of age at arrival with total fertility rate in the country of origin. If, on average, immigrants from non-English-speaking countries come from lower or higher fertility countries than those from English-speaking countries, the interaction of age at arrival with coming from non-English speaking countries (i.e., the instrument for English language proficiency in our IV estimation) captures the compound effects of language proficiency and differential fertility behaviour in the origin country. Column (2) of Table 6 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 which the effect of English proficiency becomes insignificant, although the point estimate is not significantly different from our main result in Table 3.[11]

In a similar spirit, to account for the economic context in origin countries, that might affect differently non-language age-at-arrival effects faced by immigrants from our two groups of countries, column (3) of Table 6 controls for an interaction with origin country per capita gross domestic product (GDP). In column (4), in the regression for birthweight, we also control for an interaction with infant mortality rate in the origin country, a measure of infant health. Note that interpretations of results after controlling for per capita GDP must be interpreted with caution. Due to a large number of missing values in our data on GDP, sample sizes reduce by approximately 20 per cent in every regression. Results reported in column (3) show that English language proficiency delays the age at which a woman has her first child, lowers her likelihood of becoming a teenage mother, and reduces the number of children she has, although these estimates are imprecise (at least partly) due to a significant increase in standard errors.

Turning to health outcomes, columns (2) to (4) in Table 7 control for interactions with per capita health expenditure, life expectancy, and per capita GDP in the country of origin, respec-

[10]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 the year.

[11]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 also not sensitive to the inclusion of this additional control.

Table 6: 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	GDP x age at arrival	Infant mortality rate x age at arrival
	(1)	(2)	(3)	(4)
Age at having first child	3.689**	4.024*	6.786	
	(1.780)	(2.418)	(4.685)	
Teenage mother	-0.194**	-0.218**	-0.165	
	(0.0890)	(0.0956)	(0.160)	
Number of children	-0.704*	-0.526	-0.828	
	(0.356)	(0.421)	(0.569)	
Birthweight of child	-29.76	-17.05	-150.1	-53.76
	(99.89)	(121.2)	(234.7)	(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 2 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 (1575, 1309, 1542 for each specification shown in columns 1 to 3, respectively); Teenage mother (2701, 2239, 2636); Number of children (1844, 1536, 1807); Birthweight of child (4867, 4002, 4762, 4272, for each specification shown in columns 1 to 4).
Source: Authors' calculations based on the dataset from the Office for National Statistics England and Wales Longitudinal Study.

Table 7: 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	GDP x age at arrival
	(1)	(2)	(3)	(4)
Self-reported health	0.231	0.710	0.222	-0.0477
	(0.175)	(0.509)	(0.207)	(0.273)
Good health	0.0466	0.0542	0.0107	-0.0831
	(0.0736)	(0.199)	(0.0863)	(0.110)
Bad health	0.00688	0.126	0.0275	0.0783
	(0.0417)	(0.133)	(0.0498)	(0.0612)
Long-term health problem	0.0676	0.148	0.0821	0.128
	(0.0720)	(0.222)	(0.0842)	(0.105)

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 2 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: Per capita health expenditure in 1995 (Column 2), life expectancy in 1970 (Column 3), and GDP per capita in 1970 (Column 4). The number of observations that corresponds to each of these specifications in columns 1 to 4 varies by specification: Column 1 (8,158 observations), column 2 (6,494), column 3 (7,666), and column 4 (7,988).
Source: Authors' calculations based on the dataset from the Office for National Statistics England and Wales Longitudinal Study.

tively. Note that the figures used for per capita health expenditure are from 1995 (it is the earliest year for which data is available). Table 7 shows that our results are not sensitive to the inclusion of these additional controls, thus supporting our conclusion that proficiency in English does not have significant effects on self-reported health outcomes.

Regarding educational outcomes, columns 2 to 4 of Table 8 control for interactions with average years of schooling, pupil-teacher ratio in secondary education, and per capita GDP in the country of origin, respectively. Note that as was the case with Table 6, due to missing values in origin country characteristics, sample sizes greatly reduce by approximately 20 per cent. Our estimation results, summarised in Table 8, indicate that our results are not sensitive to the inclusion of these additional variables. A difference to be noted is that the effect of English skills on the likelihood of having academic degrees is now imprecisely estimated (row 4), although the point estimate is not significantly different from the corresponding estimate in Table 3.

Table 8: 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	GDP x age at arrival
	(1)	(2)	(3)	(4)
No qualifications	-0.521***	-0.474***	-0.382***	-0.639***
	(0.0689)	(0.113)	(0.119)	(0.201)
Compulsory	-0.167	-0.126	0.162	-0.126
	(0.123)	(0.206)	(0.283)	(0.280)
Post-compulsory	0.175	0.148	-0.139	0.130
	(0.122)	(0.204)	(0.282)	(0.273)
Academic degree	0.368***	0.294	0.0985	0.343
	(0.138)	(0.241)	(0.305)	(0.340)

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 2 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: Average number of years of education of individuals aged 25 and over (Column 2), pupil-teacher ratio in Secondary school (Column 3), and GDP per capita (Column 4). The number of observations that corresponds to each of these specifications in columns 1 to 4 varies by specification: Column 1 (8,158 observations), column 2 (6,494), column 3 (7,239), and column 4 (6,868).
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 education 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 coutry which is 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. Despite these differences, our results are similar to those obtained in studies based on the US context.

In this paper, we study how speaking English affects the integration of immigrants in England and Wales. Our findings suggest that the choices that immigrants make regarding their education and fertility behaviour 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 teenage mothers, decrease the number of children they give birth to, but have no effect on their children's birthweight, a measure of child health. We also find no effects of English skills on adult self-reported health. These results could be partly motivated by the fact that speaking the language proficiently allows immigrants to better know and integrate into the UK culture and have easier access to information on family planning and public health services, and more willingness to make use of these services. We also find that immigrants who speak English proficiently are more likely to have post-compulsory qualifications and academic degrees, as compared to immigrants not proficient in English who are more likely to have no qualifications or only compulsory-level qualifications.

Our analysis is conducted exploiting the causal effect of English language skills on 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 identify 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. This idea is based on the critical period of language acquisition hypothesis, which implies that age at arrival in the UK would affect English language proficiency of immigrant children arriving from countries where English is not spoken as a main language. To partial out age-at-arrival effects that can affect the social outcomes of immigrants through channels different from language acquisition, we incorporate immigrants from English speaking countries in our analysis. 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 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 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 UK education systems and be better informed of family planning and related resources available to England and Wales citizens. This in turn may improve their educational attainment and affect fertility choices. Second, specific English learning programs at school for young immigrants that arrived in the UK after ages 8 or 9 would help them improve their 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. This support could have an important impact in their capability to complete compulsory education and pursue higher degree studies, becoming more productive contributors to the UK economy. 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 in the education and health systems of England and Wales makes them more productive, which in turn benefits the society as a whole.

References

Adsera, A., & Ferrer, A. (2011). Age at Migration, Language and Fertility Patterns among Migrants to Canada. IZA discussion paper 5552, Institute for the Study of Labour.

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

Akbulut-Yuksel, M., Bleakley, H., & Chin, A. (2011). The effects of english proficiency among childhood immigrants: Are hispanics different? In D. L. Leal, & S. J. Trejo (Eds.) *Latinos and the Economy*, Immigrants and Minorities, Politics and Policy, (pp. 255–283). Springer New York.

Barro, R., & Lee, J.-W. (2013). A new data set of educational attainment in the world 1950–2010. *Journal of Development Economics*, *104*(C), 184–198.

Bauer, A. M., Chen, C.-N., & Alegria, M. (2012). Prevalence of physical symptoms and their association with raceethnicity and acculturation in the united states. *General Hospital Psychiatry*, *34*(4), 323 – 331.

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

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

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

Böhlmark, A. (2009). Integration of childhood immigrants in the short and long run - swedish evidence. *International Migration Review*, *43*(2), 387–409.

Bratsberg, B., Raaum, O., & Roed, K. (2011). Educating children of immigrants: Closing the gap in norwegian schools. IZA Discussion Papers 6138, Institute for the Study of Labor (IZA).

Chiswick, B. R., & DebBurman, N. (2004). Educational attainment: analysis by immigrant generation. *Economics of Education Review*, *23*(4), 361–379.

Chiswick, B. R., & Miller, P. W. (2014). International Migration and the Economics of Language. In B. R. Chiswick, & P. W. Miller (Eds.) *Handbook on the Economics of International Migration*. Elsevier.

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

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

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

Feenstra, R. C., Inklaar, R., & P.Timmer, M. (2013). *The Next Generation of the Penn World Table*. Groningen Growth and Development Centre.
URL `http://www.ggdc.net/pwt/`

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

Gonzalez, A. (2003). The education and wages of immigrant children: the impact of age at arrival. *Economics of Education Review*, *22*(2), 203–212.

Gorwaney, N., Van Arsdol, M. D., Heer, D. M., & Schuerman, L. A. (1991). 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.

Guven, C., & Islam, A. (2013). Age at migration, language proficiency and socio-economic outcomes : Evidence from australia. Discussion paper, Monash University Department of Economics.

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

Janssen, S. (2010). *The World Almanac and Book of Facts 2011*. New York: Infobase Learning.

Jayaweera, H. (2014). Health of migrants in the uk: What do we know? The migration observatory, University of Oxford. Accessed on 11 December 2014.
URL `http://migrationobservatory.ox.ac.uk/briefings/health-migrants-uk-what-do-we-know`

Kimbro, R. T., Gorman, B. K., & Schachter, A. (2012). Acculturation and self-rated health among latino and asian immigrants to the united states. *Social Problems*, *59*(3), pp. 341–363.

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

Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.

Lichter, D. T., Johnson, K. M., Turner, R. N., & Churilla, A. (2012). Hispanic assimilation and fertility in new u.s. destinations. *International Migration Review*, *46*(4), 767–791.

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

Ng, E., Pottie, K., & Spitzer, D. (2011). Official language profi ciency and self-reported health among immigrants to canada. Health Reports 4, Statistics Canada.

Ng, E., Pottie, K., Spitzer, D., Mohammed, A., & Glazier, R. (2008). 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.

OECD (2012). Settling in: Oecd indicators of immigrant integration 2012. Accessed on 11 December 2014.
URL `http://dx.doi.org/10.1787/9789264171534-en`

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

Perreira, K., Harris, K., & Lee, D. (2006). Making it in america: High school completion by immigrant and native youth. *Demography*, *43*(3), 511–536.

Portes, A., & MacLeod, D. (1999). Educating the second generation: Determinants of academic achievement among children of immigrants in the united states. *Journal of Ethnic and Migration Studies*, *25*(3), 373–396.

Schaafsma, J., & Sweetman, A. (2001). Immigrant earnings: age at immigration matters. *Canadian Journal of Economics*, *34*(4), 1066–1099.

Swicegood, G., Bean, F. D., Stephen, E. H., & Opitz, W. (1988). Language usage and fertility in the mexican-origin population of the united states. *Demography*, *25*(1), pp. 17–33.