

DISCUSSION PAPER SERIES

No. 2487

**LANGUAGE PROFICIENCY AND LABOUR
MARKET PERFORMANCE OF
IMMIGRANTS IN THE UK**

Christian Dustmann and Francesca Fabbri

LABOUR ECONOMICS



Centre for Economic Policy Research

Now available online at www.cepr.org

LANGUAGE PROFICIENCY AND LABOUR MARKET PERFORMANCE OF IMMIGRANTS IN THE UK

Christian Dustmann, University College London and CEPR
Francesca Fabbri, University College London

Discussion Paper No. 2487
June 2000

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: <http://www.cepr.org>

This Discussion Paper is issued under the auspices of the Centre's research programme in **Labour Economics**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Economic and Social Research Council, under which an ESRC Resource Centre operates within CEPR; the Esmée Fairbairn Charitable Trust; and the Bank of England. These organizations do not give prior review to the Centre's publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Christian Dustmann and Francesca Fabbri

June 2000

ABSTRACT

Language Proficiency and Labour Market Performance of Immigrants in the UK*

This Paper uses two recent UK surveys to investigate labour market performance, the determinants of language proficiency, and the effect of language on earnings and employment probabilities of non-white immigrants. Our results show that language acquisition, employment probabilities, as well as earnings differ widely across non-white immigrants, according to their ethnic origin. Language has a strong and positive effect on employment probabilities. Furthermore, lack of English fluency leads to substantial earnings losses for immigrants. While earnings of white and ethnic minority natives develop in a similar manner, there is a large earnings gap between these two groups and ethnic minority immigrants. English fluency contributes considerably to reducing these differences. Addressing the problems of measurement error and unobserved heterogeneity in language variables, our results indicate that measurement error in the language variable leads to underestimation of the importance of language for employment probabilities and earnings in straightforward regressions. In comparison with results found for other countries, language proficiency seems to be more important for labour market outcomes of UK immigrants.

JEL Classification: J15, J24, J61, R23

Keywords: economics of minorities, human capital formation, immigrant workers

Christian Dustmann
Francesca Fabbri
Department of Economics
University College London
Gower Street
London WC1E 6BT
Tel: (44 20) 7504 5212
(44 20) 7504 5237
Fax: (44 20) 7916 2775 (both)
Email: f.fabbri@ucl.ac.uk
c.dustmann@ucl.ac.uk

* We are grateful to Barry Chiswick and Ian Preston for comments on earlier drafts of this Paper. This Paper is produced as part of a CEPR research programme on Labour Demand, Education and the Dynamics of Social Exclusion, supported by a grant from the Commission of the European Communities under its Targeted Socio-Economic Research Programme (no. SOE2-CT97-3052).

Submitted 13 April 2000

NON-TECHNICAL SUMMARY

According to the 1994 Labour Force Survey, ethnic minorities account for almost 5.5% of the overall population of Britain and for 6% of its working-age population. Minorities are concentrated in the areas of Greater London and of the West Midlands, where they account for 20% and for more than 14% of the population respectively (see Sly (1995) and Green and Owen (1995) for more details).

Issues surrounding the economic performance and wellbeing of minorities have received a lot of attention in the public discussion in Britain. The question of economic assimilation has always played a major role in the immigration debate. There seems to be an understanding that economic assimilation is socially desirable and that policy-makers should support this process, either by programmes aimed at the resident migrant population, or by selection of incoming immigrants.

This Paper uses two recent UK surveys to investigate labour market performance, the determinants of language proficiency and the effect of language on earnings and employment probabilities of non-white immigrants. Whether implementation of language programmes is worthwhile depends on the size of the language effect on labour market outcomes. If it is small, the cost of language education, or screening of immigrants according to their language fluency upon immigration, may outweigh the benefits. It is therefore most important to obtain a reliable estimate of this coefficient. We attempt to address the problems, which arise in statistical analyses when estimating the effects of language on labour market outcomes.

Our results show that language acquisition, employment probabilities, as well as earnings differ widely across non-white immigrants, according to their ethnic origin. We find that in simple regressions, language proficiency is strongly associated with higher employment probabilities and with higher earnings. While earnings of white and ethnic minority natives develop in a similar manner, there is a large earnings gap between these two groups and ethnic minority immigrants. On average, the earnings disadvantage does not disappear over time. English fluency contributes considerably to reducing these differences.

Some simple back of the envelope calculations suggest that an immigrant, who arrives at age 20 and who does not acquire fluency in English over the first ten years of his/her migration history, loses on average between £10,000 and £21,000, compared to an immigrant who arrives with language fluency (depending on the data source). Accordingly, language proficiency is an important factor for economic success of immigrants in the UK and the

earnings losses immigrants experience due to a lack of language fluency are substantial.

There are also large differences in employment probabilities and earnings between individuals who belong to different minority communities. Immigrants from Bangladesh are the most disadvantaged group. Chinese immigrants, the most successful group, have about 50% higher earnings than Bangladeshi immigrants. Similar patterns of ethnic differences in earnings can be observed between ethnic minority immigrants and ethnic minority natives.

Our results also indicate that measurement error in the language variable leads to underestimation of the importance of language for employment probabilities and earnings in straightforward regressions.

We compare estimates obtained for the UK with those found for Germany, the US, Canada, Australia and Israel. The association between language and earnings is largest for the UK, followed by the US and Canada. Much smaller effects are found for Germany, Australia and Israel. Accordingly, language seems to be more significant in affecting earnings of UK immigrants than those of immigrants to other countries. As a consequence, it may be worthwhile to implement schooling centres that support immigrants in their acquisition of the English language at an early stage of their migration history. Given the substantial earnings disadvantages immigrants experience due to a lack of English fluency, there is clearly an incentive for the migrants to bear part of the costs of language education.

1 Introduction

According to the 1994 Labour Force Survey, ethnic minorities account for almost 5.5 per cent of the overall population of Britain, and for 6 per cent of its working-age population. Minorities are concentrated in the areas of Greater London and of the West Midlands, where they account for 20% and for more than 14% of the population respectively (see Sly (1995) and Green and Owen (1995) for more details).

Issues surrounding the economic performance and wellbeing of minorities have received a lot of attention in public discussion in Britain. The question of economic assimilation has always played a major role in the immigration debate. There seems to be an understanding that economic assimilation is socially desirable, and policy makers should support this process, either by programmes aimed at the resident migrant population, or by selection of incoming immigrants.

A number of recent studies analyse various aspects of labour market behaviour of ethnic minorities, and compare outcomes with those of the majority population (see e.g. Blackaby et al. (1994, 1997)). In much of this literature, however, no attempt is made to distinguish between immigrant and British born minorities. But many important questions are specifically related to first generation immigrants, who constitute a significant fraction of minorities in the UK. Out of a total of 2.6 million individuals belonging to ethnic minorities, over half are foreign born (Sly (1995)). This percentage is even higher when considering ethnic minorities of

working age, where 73% are born abroad (Sly (1995)). We show in this paper that native born and migrant minority populations differ quite substantially in terms of their economic success, and that distinguishing the two groups is important for addressing identification of economic differences.

Substantial initial earnings disadvantages of immigrants, as observed in a number of studies for the US (see, e.g. Chiswick (1978), Borjas (1985)), are often explained by migrants having lower levels of human capital when arriving in the host economy. The process of economic assimilation of immigrants depends then on the intensity with which they acquire host country specific skills. For the UK, the economic assimilation of immigrants has been analysed in papers by Chiswick (1980), Bell (1997), and Shields and Wheatley-Price (1998)). Chiswick (1980) uses data from the 1972 GHS. His main finding is that, while white immigrants have very similar earnings pattern to native-born individuals, earnings of coloured immigrants are about 25 percent lower, other things the same. This gap is not decreasing with time of residence in the UK. In a more recent paper, Bell (1997) uses also data from the General Household Survey, but he pools waves between 1973 to 1992. Like Chiswick, he finds that white immigrants are doing surprisingly well. While white immigrants have an initial wage advantage, compared to native workers, black immigrants have an earnings disadvantage, but wage differentials between this group and white natives decrease with the time spent in the UK. Shields and Wheatley-Price (1998) use data from the British Labour Force Survey. They emphasise the

different assimilation pattern between foreign and native born minority individuals.

It may be in the interest of the host country to support the process of economic assimilation. To achieve this, it is important to understand the factors which determine the economic performance of minority immigrants. In this paper, we concentrate on one specific human capital factor, which is important not only for immigrants' economic assimilation, but also for their social integration: Language. Recent analyses for the US, Canada, Australia, Israel, and Germany show that fluency and literacy in the dominant host country language are important components for explaining immigrants' labour market success (see, e.g., Rivera-Batiz (1990), Chiswick (1991), Chiswick and Miller (1995), Chiswick, Cohen and Zach (1997), Dustmann (1994)). Work by Shields and Wheatley-Price (1999) indicates that language is also positively related to occupational success of some immigrant groups in the UK.

Enforcement and support of language acquisition is a possible policy instrument. If the returns to language capital are sufficiently large, governments may find it worthwhile to invest into programmes which enhance language skills of immigrants. Whether implementation of language programmes is worthwhile depends on the size of the language effect on labour market outcomes. If it is small, the cost of language education, or screening of immigrants according to their language fluency upon immigration, may outweigh the benefits. It is therefore most important to

obtain a reliable estimate of this coefficient.

Some governments find it worthwhile to require, or to enforce language proficiency for their immigrant population. Some countries have implemented very rigorous language tests for immigrant applicants (e.g. New Zealand, which even requires the partner of the applicant to be proficient), while others require immigrants to undergo an intensive language education upon entry to the host country (e.g. Israel; see Eckstein and Weiss (1999) for details). So far, most European countries have been reluctant to implement policies which impose entry conditions requiring an array of skills of immigrant applicants, or which demand the acquisition of skills upon entry.

In this paper, we analyse the determinants of fluency and literacy in the host language for immigrants belonging to ethnic minority groups, and on how it relates to their labour market performance. We first investigate factors influencing the acquisition of the host country's language by the immigrant, such as education, age, and years of residence in the host country. We distinguish between education received in the host- and in the home countries.

We then analyse the extent to which language ability influences labour market outcomes of immigrants. We focus on its effect on employment probabilities, and on the level of earnings. We compare earnings paths of minority immigrants with ethnic minority and white native born individuals. Our results show similarities

between native born whites and ethnic minority individuals, but dramatic differences between these two groups, and foreign born minorities. Language is a most important determinant in explaining earnings differences among these groups.

As the discussion indicates, it is most important to obtain a precise estimate of the effect of language on labour market outcomes. Unfortunately, estimates of language coefficients in straightforward regressions are bedeviled by two problems. First, as pointed out by Borjas (1994), language may be correlated with unobserved ability components which also affect the outcome variable. Second, as stressed by Dustmann and vanSoest (1998), language measures usually reported in survey data may suffer substantially from measurement error. The bias induced by these two problems points in opposite directions. We attempt to address both problems in this paper, and propose estimators which may help to reduce, or eliminate the bias.

We base our analysis on data from two UK surveys on ethnic minorities: the Fourth National Survey on Ethnic Minorities (FNSEM), which has been collected between 1993 and 1994, and the Family and Working Lives Survey (FWLS), which has been collected between 1994 and 1995. Both data sets consist of two subsamples. The FWLS contains a main sample of the entire UK population, and a "boost" sample of individuals belonging to ethnic minorities. The FNSEM contains a main sample of respondents belonging to ethnic minorities, and a reference sample of individuals belonging to the white majority population. Both surveys include questions

on social and economic conditions of the interviewees, and measures on language proficiency. Information in the two data sets is complementary. For instance, while the FNSEM only reports spoken language proficiency, the FWLS contains also information about reading and writing skills. Also, the FNSEM distinguishes between education acquired in home- and host economy, which is not available for the FWLS. Using two datasets allows us to conduct comparable analyses to check the robustness of the results obtained.

The structure of the paper is as follows. Section 2 develops the estimation equations. Section 3 briefly describes the data sets, and gives some descriptive statistics. Section 4 investigates language determinants. Section 5 analyses how language proficiency affects the individual's probability of being employed, and earnings, and draws comparisons between ethnic minority immigrants, and native white and ethnic minority individuals. Section 6 summarises the results obtained, and compares findings for the UK with those for other countries.

2 Language and Labour Market Outcomes

The literature on migrants' earnings assimilation distinguishes between human capital which is specific to the host country, human capital which is specific to the home country, and human capital which is equally productive in both countries. Typically, immigrants enter the host country with skills which are only of limited use in

the host economy, which results in an initial earnings disadvantage (see Chiswick (1978)). After immigration, migrants transfer home country specific human capital into general or host country specific human capital, and acquire additional skills which are specific to the host country economy. The intensity of this process determines the speed of economic assimilation.

Language capital is a most important component of host country human capital. Furthermore, it is very specific to the host economy, since it is usually not transferable to the migrant's home economy. Standard human capital models may serve as a basis to formulate empirical specifications explaining the determinants of language capital. In such models, human capital is produced by investing time and other inputs. The cost of production equals forgone earnings, plus the cost of other input goods. A simple equilibrium condition states that investment into human capital production is set such that the cost equals the benefit from the discounted future enhanced earnings potential. The production potential may differ across individuals according to their ability to acquire knowledge, and it may depend on the stock of human capital acquired in the past. The benefit of any acquisition of host country specific human capital depends, in addition, on the length of the period over which it is productively put into use. In the case of language capital of immigrants, this corresponds to the migration period.

Accordingly, investment into language capital should depend on its potential

future benefits, on the cost of acquisition, and on the individual's efficiency in producing it. Furthermore, it should be affected by the envisaged migration period. Variables which measure the immigrant's efficiency in acquiring language capital are the level of education upon immigration, and the age at immigration (since the learning potential may deteriorate over the life cycle). The cost of acquiring the host country language depends on the distance of the migrant's mother tongue to the dominant majority language, which may be captured by country of origin dummies. Clearly, this last variable picks up a variety of other factors which affect language proficiency, like different degrees of immigrant selection across countries (see Borjas (1985, 1987)). Assuming that all migrations are permanent, the time period over which any language capital is productive depends on the migrant's age at entry. Accordingly, those who migrate at younger age should have a higher incentive to acquire language capital. The acquisition of language capital may, in addition, depend on the extent to which individuals are exposed to the language of the majority population. A variable which measures exposure is the time of residence abroad.

Finally, it is likely that the value of language capital differs across locations in the host country, according to the relative size of the ethnic minority population the immigrant belongs to. Lazear (1999) develops a model where in each period individuals encounter each other and get involved into trade activities. Trade can only take place between individuals who have the same means of communication (language, for instance). The smaller the size of the minority population, the smaller is the

probability that an individual of that population who is not fluent in the native language will get involved into successful trading without mastering the host language. Accordingly, given the cost of language acquisition, the smaller the relative size of the minority population, the larger will be the probability that an individual acquires the native language. The empirical implication of this is that immigrants in areas of high ethnic concentration should be less proficient in the host language.

We summarise the factors which explain language acquisition in the vector of observed variables z_i , where i is an index for the i th individual. The language equation can be written as

$$l_i = z_i' \delta + u_i, \tag{1}$$

where u_i is an error term. Assumptions about u_i determine the nature of the model and the properties of its estimator.

2.1 Measurement Error and Unobserved Heterogeneity

A second objective of the paper is to analyse the effect of language on labour market outcomes, in particular on wages and employment probabilities. Dustmann and van Soest (1998) discuss in detail the problems which may occur in straightforward regression analysis, where language is a regressor. First, the same unobserved heterogeneity components may affect both language (as discussed above), and economic

performance, leading to an upward bias of estimated language effects on economic outcomes. Second, unsystematic measurement error may lead to a downward bias of the effect of language on earnings. Numbers presented in Dustmann and van Soest (1998) on repeated language information for the same individual suggest that measurement error is substantial in self-reported language measures. In fact, in their data, more than half of the within individual variation in language responses is due to measurement error. Their results suggest that the downward bias induced by measurement error overcompensates the upward bias induced by unobserved heterogeneity.

To explain the issues involved in more detail, consider a generic outcome function, where y_i is log earnings or (the underlying index of) the employment equation respectively:

$$y_i = x_i' \beta + \gamma l_i + \alpha_i + v_i, \quad (2)$$

where x_i are (exogenous) variables which determine the outcome variable, l_i is the ‘true’ language proficiency of the individual, and i is an index for individuals, respectively. The error term α_i is unobserved individual heterogeneity, while v_i is an idiosyncratic error term.

In the general model, we do not observe l_i , but only \tilde{l}_i . We assume that $\tilde{l}_i = l_i + \eta_i$, where η_i denotes an unsystematic measurement error.

Substitution in (2) gives

$$y_i = x_i' \beta + \gamma \tilde{l}_i + \alpha_i + v_i - \gamma \eta_i. \quad (3)$$

Throughout, we assume that α_i , η_i , and v_i are independent of each other and of the x_i , and that the measurement error η_i is uncorrelated with l_i .

OLS on equation (3) leads to inconsistent estimates of γ if $E(\alpha_i + v_i - \gamma \eta_i | \tilde{l}_i) \neq 0$. This will generally be the case if there is measurement error, or if l_i and the individual specific heterogeneity α_i are correlated.

The asymptotic bias of the OLS estimator is given by

$$\text{plim}(\hat{\gamma}_{OLS} - \gamma) = \frac{\sigma_{l\alpha} - \gamma \sigma_\eta^2}{\sigma_{\tilde{l}}^2 (1 - R_{\tilde{l}x}^2)}, \quad (4)$$

where $\sigma_{l\alpha}$ is the covariance between α_i and l_i , and σ_η^2 the variance of the measurement error. Furthermore, $\sigma_{\tilde{l}}^2$ is the variance of \tilde{l}_i , and $R_{\tilde{l}x}$ is the multiple correlation coefficient in a regression of \tilde{l} on x . The first term in the numerator is the bias due to unobserved heterogeneity. It is positive if the same unobserved component affects earnings and language proficiency in the same direction. The second term is the asymptotic (downward) bias due to measurement error.

To give a structural interpretation to the language coefficient, we need to deal with both sources of bias. We address the two issues separately. To deal with the

measurement error, we use IV estimation. As instruments, we use information on the language in which the survey interview has been conducted, information which is available in one of our datasets (the FNSEM). These instruments are however still correlated with unobserved heterogeneity components. To reduce the bias due to unobserved heterogeneity, we add partner variables and household characteristics to the earnings equation. These variables are likely to be correlated with the unobserved heterogeneity component in language and outcome equation. Inclusion of these background characteristics may therefore reduce the correlation between α_i and l_{it} , thus reducing the bias. The resulting OLS estimator can be interpreted as a matching estimator in the sense of Rosenbaum and Rubin (1983). If, conditional on these background characteristics, l_i is uncorrelated with α_i , this procedure eliminates the bias due to unobserved heterogeneity.

3 The Data

The Family and Working Lives Survey (FWLS) has been collected in 1994 and 1995. It is a retrospective survey on adults aged between 16 and 69, including 9000 respondents and their partners. It contains a "boost" sample of about 2000 individuals belonging to four racial minority groups: Black Caribbeans, Indians, Pakistanis and Bangladeshis. The data provides information on earnings, education, nationality, language skills and parents' economic and educational status. Of the

2388 people forming the minority sample in the main and "boost" sample, 68% (1639) are foreign born.

The Fourth National Survey on Ethnic Minorities (FNSEM) is also a cross-sectional survey, which has been carried out between 1993 and 1994. Individuals included are aged 16 or more, and of Caribbean, Indian, Pakistani, Bangladeshi, or Chinese origin. There are 5196 observations in the minority sample, and 2867 observations in the independent comparison sample of white individuals. Similarly to the FWLS, more than 77% (4019) of the individuals in the ethnic minority sample are foreign born.

The FWLS identifies the ward where the individual lives. It is therefore possible to match this data set with the 1991 Population Census to construct a variable on the ethnic concentration on ward level. The FNSEM does not contain geographical identifiers; therefore, matching with the Census data is not possible. It contains however grouped information on ethnic concentration at ward level.

Both data sets provide information on earnings. The FWLS reports gross (before tax) earnings, while the FNSEM reports grouped gross weekly earnings. Both data sets report the main activity of the individual (e.g. full-time or part-time paid work, full-time education, unemployed, etc.).

The sample design of the two surveys differs substantially. The ethnic minority sample of the FWLS was selected by screening addresses in areas where the

ethnic minority population, according to the 1991 census, was more than 3% of the local population. The selection in the FNSEM was more complex, considering wards with any percentage of ethnic minorities on the population and oversampling Bangladeshis to obtain a sufficient sample size. For more details, see Appendix 1 in Modood *et al.* (1997), and Smith and Prior (1996).

Table 1 shows the percentage of immigrants belonging to ethnic minorities with respect to the overall population in the UK (column 1), and the ethnic composition within the group of ethnic immigrants. Numbers are based on the 1991 Census. Table 2 gives the ethnic composition of the two surveys. Both surveys do not include Black African immigrants, and the FWLS does not include the Chinese minority. In the last column of table 1, we report respective numbers in the census, excluding Africans. Comparing the two tables, it appears that both surveys tend to oversample the South Asian groups (Indians, Pakistanis and Bangladeshis). Also, the two surveys differ in the ethnic composition of the respondents: Bangladeshis amount to 31% in the FWLS and 14% in the FNSEM, Indians to 19% in the FWLS and 24% in the FNSEM and African Asians to 8% in the FWLS and 17% in the FNSEM.

Table 1: Ethnic Immigrants Composition in the UK (Census 1991)

	Immigrants Perc. wrt UK Pop.	Ethnic composition	Ethnic composition without Africans
Caribbean	0.56	18.19	23.41
Indian	0.84	27.57	35.49
African	0.68	22.31	-
Bangladeshi	0.22	7.09	9.13
Pakistani	0.47	15.46	18.89
South East Asians	0.29	9.37	12.06
Total	3.06	100	100

Table 2: Ethnic Composition

Variable	FWLS		FNSEM		
	No.	Perc.	No.	Perc.	Perc.
Black Caribbeans	265	16.17	698	18.20	17.37
Indians	314	19.16	971	25.32	24.17
Afro-Asian	123	7.50	656	17.11	16.32
Pakistanis	425	25.93	960	25.05	23.89
Bangladeshis	512	31.24	550	14.34	13.68
Chinese	-	-	184	-	4.58
Total	1639	100	4019	100	100

Both surveys contain information on language. In the FWLS, language ability is self-assessed. The individual is first asked whether s/he speaks English as mother tongue. If not, the individual is asked to self-assess proficiency in speaking, reading, and writing English on a 5 point scale. The FNSEM contains two variables which are related to language proficiency: first, the interviewer's evaluation on the individual's spoken language ability, on a 4 point scale. Second, information about what fraction of the interview was held in English.

In Table A1 we display the complete responses to the language questions for the

two data sets, broken down according to ethnic origin. The general pattern is similar for the two data sets. In fact, the percentage of individuals fluent or very fluent in English in the FNSEM equals the percentage of individuals speaking English as first language plus the percentage of individuals fluent or very fluent in the FWLS.

For the empirical analysis, we re-define the language indicators in the two surveys to dichotomous variables. For the FWLS, this variable assumes the value 1 if the individual reports language fluency or literacy as "well" or "very well", or reports English as a first Language. For the FNSEM, it is equal to 1 if individuals fall in the categories "fairly well" or "fluently".

Table 3 explains the variables used for the analysis, and presents summary statistics. The mean values on language indicate that the percentage of individuals who speaks the English language well or very well is very similar in the two samples. Percentages for reading and writing in English (available in the FWLS) are slightly lower.

Table 3: Variables Description and Sample Characteristics

Variable	FWLS		FNSEM		Description
	Mean	S.D.	Mean	S.D.	
Speak	0.709	0.454	0.691	0.462	Dummy=1 if spoken English is good or very good
Read	0.671	0.469	-	-	Dummy=1 if read English is good or very good
Write	0.641	0.479	-	-	Dummy=1 if written English is good or very good
Empl	0.367	0.482	0.425	0.494	Dummy=1 if employed
Unempl	0.162	0.369	0.138	0.345	Dummy=1 if unemployed
Wgearn	239.175	432.809	221.185	-	Weekly gross earnings
Sex	0.468	0.499	0.505	0.500	Dummy=1 if male
Age	38.347	13.588	42.604	14.407	Age
Yearstay	20.404	10.313	21.367	10.001	Years of stay in the UK
Married	0.726	0.446	0.776	0.417	Dummy=1 if married
Nchild	1.937	1.793	1.654	1.761	Number of children in household
Degree	0.072	0.258	0.127	0.333	Dummy=1 if university degree
Alev	0.129	0.335	0.109	0.312	Dummy=1 if high vocational
Voc	0.231	0.422	0.230	0.421	Dummy=1 if low vocational
Noqual	0.568	0.495	0.533	0.499	Dummy=1 if no qualification
Immcon	0.090	0.094	-	-	Ward own ethnic immigrants concentration
Ethcon	0.168	0.153	0.166	0.189	Ward own ethnic concentration
Carib	0.1620	0.369	0.178	0.383	Dummy=1 if Black Caribbean
Indian	0.1860	0.389	0.245	0.429	Dummy=1 if Indian
Afroas	0.0838	0.277	0.169	0.375	Dummy=1 if African Asian
Pakista	0.255	0.436	0.218	0.413	Dummy=1 if Pakistani
Chinese	-	-	0.048	0.214	Dummy=1 if Chinese
Bangla	0.318	0.466	0.142	0.349	Dummy=1 if Bangladeshi

About 40% of the sample populations are employed, and 33% out of the labour force. These numbers are remarkably similar for the two data sets. Furthermore, 16% (FWLS) and 14% (FNSEM) report being unemployed.

The mean value of weekly wages in the FWLS is £239.17, considering both part and full-time workers. Mean weekly wage are reported in the FNSEM as a grouped variable. The mean weekly gross wage is 221 £, which is similar to the mean wage

in the FWLS.¹

The average education level is slightly higher in the FNSEM than in the FWLS, with 12.7% graduates in the former sample, and only 7.2% in the latter sample. Furthermore, there is a slightly higher percentage of individuals with no qualification in the FWLS (56.8%) than in the FNSEM (53.3%).²

The average ethnic minority concentration at ward level amounts, in both samples, to more than 16% (the average ward concentration in the FNSEM is obtained by taking the average of the mid-point values of the grouped variable, since the information is available only in intervals). The considerable difference in the sample designs is reflected only by the larger standard deviation indicated in the FNSEM.

In Table A2, we break down means of the age at immigration, year of immigration, and age for the various ethnic groups. In the FWLS, individuals are on average four years younger than in the FNSEM, and have immigrated at a younger age. The immigration patterns for the various ethnic groups are similar in both data sets, and correspond to the migration patterns indicated by Bell (1997) and Hutton and Wheatley Price (1999): Black Caribbeans arrivals are concentrated in the late 1950's and early 1960's, whereas Indians, African Asians and Pakistanis arrived mainly during the 1970's, and Bangladeshis towards the end of the 1970's.

¹To obtain this number, we estimate a grouped regression model (see Steward, 1983) on a constant, and compute the mean of the predictions.

²We construct the education variables following a classification by Dearden (1999).

Consistent with their shorter stay, Bangladeshis are the youngest group, whereas Black Caribbeans are the oldest on average.

4 Language Determinants

After eliminating all the observations with missing values in the variables of interest, we are left with 1475 observations in the FWLS sample, and 3732 observations in the FNSEM sample.

Table 4 reports results from straightforward probit regressions, where the indicator variable equals one if the individual is proficient in the respective language component. We report marginal effects, evaluated at average sample characteristics. Comparing results on spoken language for the two data sets shows that the signs of regressors are equal for both samples in most cases, and the sizes of the coefficients are likewise similar (although the coding of the fluency variables differs slightly). Females have a significantly lower probability to be fluent in the majority language. The effect of age (which corresponds to the effect of age at entry, since we condition on years of residence) is negative and strongly significant, and the years of residence variable has the expected positive effect. All these results are consistent with findings for other countries. Furthermore, for the FWLS, the effect of these variables is similar for all three components of language capital.

The effect of the education variables is quite strong for fluency (the comparison

group are individuals who report to have no qualification): For the FWLS (FNSEM) individuals with vocational training have a 23 (17) percent higher probability of being fluent in English. The association between fluency and higher educational degrees is not stronger, though.

Speaking fluency may largely be acquired by exposure to the host country language, while writing and reading in a foreign language is a skill which is more difficult to obtain. Acquisition requires a more systematic way of learning, and the general level of schooling obtained may enhance the efficiency of acquiring this component of language capital. This is reflected by our results, which indicate that educational background variables have larger coefficients for reading and writing skills.³

³The variable "degree" predicts outcomes perfectly. Individuals with degrees do therefore not contribute to the likelihood, since $\text{Prob}(\text{Fluent}) = \mathbf{1}(\text{Degree}=1) + \mathbf{1}(\text{Degree}=0)\text{Prob}(z'_i\delta > -u_i)$. $\mathbf{1}(\cdot)$ is an indicator function, which does not depend on the parameter vector δ . Estimations are therefore performed on the sample of non-degree holders.

Table 4: Language determinants, Probit Estimation

Variable	FWLS						FNSEM			
	Speaking		Reading		Writing		Speaking			
					All Qualifications		UK/nonUK		Q	
	ME	t-ratio	ME	t-ratio	ME	t-ratio	ME	t-ratio	ME	t-ratio
sex	0.109	4.95	0.138	5.06	0.103	3.85	0.148	10.19	0.139	9.85
age	-0.006	-6.10	-0.008	-6.38	-0.007	-5.34	-0.012	-18.30	-0.010	-15.85
yearstay	0.008	5.40	0.009	4.76	0.007	3.71	0.013	14.07	0.011	11.81
degree	0.185	5.93	–	–	0.286	7.64	0.221	16.12	–	–
Alevtea	0.208	8.24	0.301	9.35	0.309	10.41	0.178	11.38	–	–
OlevCSE	0.226	10.98	0.308	11.85	0.316	13.12	0.172	12.78	–	–
Edegree	–	–	–	–	–	–	–	–	–	–
EAlevtea	–	–	–	–	–	–	–	–	0.188	6.34
EOlevCSE	–	–	–	–	–	–	–	–	0.209	9.56
Fdegree	–	–	–	–	–	–	–	–	0.177	13.40
FAlevtea	–	–	–	–	–	–	–	–	0.117	5.74
FOlevCSE	–	–	–	–	–	–	–	–	0.114	8.00
married	-0.067	-2.61	-0.090	-2.75	-0.073	-2.27	-0.039	-2.07	-0.012	-0.66
nchild	-0.010	-1.67	-0.009	-1.29	-0.020	-2.65	-0.005	-1.34	-0.005	-1.25
indian	0.164	6.75	0.202	6.31	0.192	5.92	0.064	3.16	0.072	3.75
afroas	0.181	5.75	0.235	5.57	0.203	4.84	0.176	10.09	0.165	9.93
pakista	0.093	4.10	0.061	1.99	0.065	2.14	-0.025	-1.23	-0.009	-0.45
carib	0.244	8.55	0.346	9.45	0.352	10.11	0.309	13.77	0.297	13.89
chinese	–	–	–	–	–	–	0.042	1.33	0.007	0.22
ethcon	-0.467	-4.68	-0.402	-3.23	-0.201	-1.63	-0.198	-5.75	-0.209	-6.15
No. of Obs.	1589		1475		1589		3732		3552	
Obs. Prob.	0.710		0.646		0.641		0.691		0.675	
Pred. Prob.	0.835		0.757		0.757		0.841		0.852	

ME: Marginal Effects, evaluated at sample means.

Excluded categories: no education qualification, Bangladeshi.

Ethnic concentration for FNSEM at mid-point

Education may be partly obtained in the host country. Since those who wish to enter the educational system in the UK are likely to have acquired some language skills, this leads to a classical simultaneity bias.

The FNSEM allows distinguishing between education obtained in the UK and abroad. We have re-estimated the language equation, distinguishing between edu-

cation obtained overseas, and in the UK. Results are reported in the last column of table 4. We denote by F educational achievements obtained abroad, and by E educational achievements obtained in the UK.⁴ The effect of overseas qualifications on language fluency is still positive, but slightly lower than the effect of education obtained in the UK. These differences in the coefficients are statistically significant at the 1 percent level.

The variable NCHILD measures the number of children in the household. Chiswick and Miller (1995) suggest that children may have counteracting effects on language: first, they may act as a translator between the parent and the English environment (thus reducing incentives to learn the foreign language). Second, they may enhance exposure to the majority population by forcing the parent to cope with institutional matters, like school and parents of native friends of children. Our results indicate that children coefficients are negative for both data sets, and for all language components.

There are large differences in the level of language proficiency among different ethnic groups. Results of both data sets indicate that Bangladeshis, the excluded group, are dominated by nearly all other ethnic groups, except for Pakistanis in the FNSEM.

The variable *ethcon* measures ethnic concentration of the own minority at ward

⁴Since "Edegree" predicts outcomes perfectly, estimation is performed on those who do not hold a UK degree. See footnote 2.

level. It is strongly associated with language proficiency for both data sets. Results from the FWLS indicate that an increase in the ethnic density by 1 percentage point is associated with a 0.47 percent decrease in the probability to be fluent in the dominant language. The negative association with reading and writing skills is slightly smaller. Results from the FNSEM also indicate a negative association, but the size of the coefficient is only half as large as that for the FWLS. Similar results are found for the US, Canada and Israel (see Chiswick (1994), and Chiswick and Miller (1995)).

Clearly, if individuals are free to choose their location, the ethnic density coefficient measures a compound effect, consisting of the direct effect (which corresponds to the effect of ethnic concentration on language when assigning individuals randomly to areas with different concentrations), and an indirect effect (which results from individuals sorting to neighbourhoods with different concentration, according to their potential to learn the language). As Lazear (1999) points out, both are consistent with his model of language being a means of decreasing the cost of trading.

However, for many policy related questions (e.g. settlement policies of arriving immigrants) it may be interesting to disentangle these two effects. To identify the causal effect of ethnic concentration is difficult, since it requires instruments which affect location choice, but not language acquisition. In a companion paper (Dustmann, Fabbri and Preston (2000)), we develop a model which shows that

the bias due to sorting may go in both directions, and we suggest identification strategies.

5 Language and Economic Outcomes

5.1 Employment Probabilities

Language proficiency is likely to be a decisive factor in determining employment probabilities. Language may help to acquire information about optimal job search strategies. Migrants who are not sufficiently proficient in the dominant language may have difficulties to convince prospective employers of their qualifications. Also, many jobs, for instance in the service sector, require communicative skills; likewise, literacy in the dominant language is a crucial prerequisite for many unskilled occupations.

To understand the association between employment probabilities and language, we consider individuals who are in the labour force, and we distinguish between those who are in work, and those who are not employed, but who are actively seeking a job.⁵ Our samples consist of 849 individuals for the FWLS, and 2165 individuals for the FNSEM. Our dependent variable, EMPL, takes the value 0 if the individual is

⁵This follows the ILO definition of unemployment. According to the ILO definition, people are considered as unemployed if aged 15 years or older, who are without work, but available to start within the next two weeks, and who have actively sought employment at some time during the previous four weeks.

unemployed and seeking a job or claiming benefits, and the value 1 if the individual works full- or part-time. Explanatory variables are the demographic and human capital characteristics available in the two data sets, including a dummy variable for the level of language proficiency.

The results are reported in Table 5. We estimate probit models, and report marginal effects, evaluated at the mean vector of sample characteristics. For the FWLS, we report results conditioning on fluency only, and on fluency and written literacy.

Most coefficient estimates for the two data sets are very similar. Females have a significantly higher probability of being employed (14 percent in the FWLS, and 10 percent in the FNSEM). Being married increases employment probabilities by about 20 (23) percentage points. Having children influences, on the other side, the employment probability negatively. These effects are consistent with evidence for British natives. Nickell (1980) finds that married individuals are about 30% more likely to being employed than unmarried ones. His interpretation for this findings is that married individuals may have more responsibilities towards their families, and have a more stable working record, both increasing their employment probabilities.

For the FWLS, education coefficients are mostly insignificant. For the FNSEM, education coefficients are significant, and in the expected order of magnitude. In the last columns of table 5, we run regressions which distinguish between education

levels acquired in the UK, and in the home country. The coefficients on the UK educational degrees seem larger than the coefficients on education acquired at home. However, only for A-levels can we reject the null hypothesis that the coefficients are equal (and only at the 10 percent level).

Table 5: Employment probabilities

Variable	FWLS						FNSEM				
	1		2		3		4				
							All Qualifications		UK/nonUK Q		
	Coeff	t-ratio	Coeff	t-ratio	Coeff	t-ratio	Coeff	t-ratio	Coeff	t-ratio	
sex	-0.143	-3.81	-0.138	-3.69	-0.140	-3.74	-0.102	-4.69	-0.100	-4.62	
married	0.207	4.32	0.208	4.33	0.210	4.36	0.229	7.40	0.231	7.48	
nchild	-0.035	-2.93	-0.034	-2.82	-0.034	-2.83	-0.025	-4.05	-0.025	-4.05	
degree	0.072	1.19	0.043	0.68	0.043	0.67	0.112	4.08	-	-	
Alevtea	0.004	0.09	-0.020	-0.40	-0.022	-0.43	0.124	4.41	-	-	
OlevCSE	-0.071	-1.65	-0.093	-2.07	-0.095	-2.10	0.073	3.31	-	-	
Edegree	-	-	-	-	-	-	-	-	0.112	3.00	
Elevtea	-	-	-	-	-	-	-	-	0.139	4.20	
ElevCSE	-	-	-	-	-	-	-	-	0.068	2.61	
Fdegree	-	-	-	-	-	-	-	-	0.090	2.69	
Flevtea	-	-	-	-	-	-	-	-	0.065	1.57	
FlevCSE	-	-	-	-	-	-	-	-	0.063	2.37	
age	0.028	2.81	0.029	2.94	0.029	2.95	0.013	2.34	0.013	2.24	
agesq/100	-0.038	-3.15	-0.040	-3.30	-0.040	-3.29	-0.021	-3.08	-0.020	-2.98	
yearstay	0.003	1.31	0.003	1.43	0.003	1.32	0.003	2.33	0.002	1.88	
black	0.095	1.61	0.083	1.39	0.080	1.32	0.078	2.18	0.077	2.14	
afroas	0.117	2.05	0.120	2.13	0.115	2.00	0.146	4.68	0.144	4.59	
indian	0.165	3.41	0.166	3.46	0.161	3.31	0.118	3.85	0.119	3.90	
pakista	0.055	1.22	0.061	1.38	0.057	1.27	0.006	0.21	0.009	0.32	
chinese	-	-	-	-	-	-	0.179	4.73	0.175	4.55	
speak	0.152	3.01	-	-	0.048	0.73	0.172	5.99	0.168	5.86	
write	-	-	0.173	3.63	0.140	2.19	-	-	-	-	
N. of Obs.	839		839		839		2100		2100		
	Excluded categories: no education qualification, Bangladeshi.										

Age is positively associated with employment probabilities, and the age profile is

concave. The time of residence in the UK has a positive effect on employment probabilities, but it is small, and not significant for the FWLS. Indians, Afro-Asians and Chinese have higher probabilities of being employed than Pakistanis and Bangladeshis. Again, Bangladeshis seem to be the most disadvantaged group.

The coefficients on the language variables are quite large. English fluency is associated with a 15 percent (17 percent) higher employment probability, using the FWLS (FNSEM) data. The coefficients are highly significant, and coefficients are remarkably similar for the two data sets.

The FWLS data distinguishes between speaking, writing and reading abilities – information which is not available in most datasets on migrants’ language abilities. One may argue that proficiency in the spoken language alone is not sufficient to affect labour market outcomes, but that writing skills are likewise needed. The positive coefficient of the fluency variable may then simply reflect the correlation between these two components of language capital. To investigate this point, we have included an indicator for writing abilities (column 2), and both speaking and writing variables (column 3). The effect of writing proficiency (unconditional on fluency) is slightly higher. When including both indicator variables, we find that writing abilities are associated with a 14 percent increase in employment probabilities, while speaking ability alone increases this probability by only 4 percentage points. The latter effect is not significant. This suggests that literacy in the dominant majority

language, in addition to fluency, is important to obtain a job.

5.2 Earnings

We now turn to analysing the effect of language on earnings. Both samples do not provide information on the number of hours worked per week. We consider here only individuals who are working full-time.

In the FWLS, the dependent variable is the natural logarithm of gross (before tax) weekly earnings. The earnings variable in the FNSEM is gross weekly earnings, which is reported in categorical form (16 categories). In both samples there is a considerable percentage of working individuals who do not report their earnings (28% in the FNSEM and 45% in the FWLS).

To check the extent to which attrition is non-random, we compare the means of the language variables, origin dummies, and the educational variables for individuals who do, and who do not report earnings. Results are presented in table A3. We also report the t-statistics for testing whether the means of the variables are significantly different. In some cases, we reject the null hypothesis of equal means, but there seems to be no systematic pattern of attrition across the two data sets.

Our final sample sizes are 259 individuals for the FWLS data, and 967 individuals for the FNSEM data. Results of straightforward log wage regressions are presented in Table 6, where we use the least squares estimator for the FWLS, and a grouped

regression model for the FNSEM (where the boundaries are transformed by taking logs).

As regressors, we include demographic and human capital characteristics, dummies for English proficiency, and dummies for ethnic background. Coefficient estimates on most variables are roughly similar for the two data sets. Males have a significant earnings advantage, compared to females. Having a degree increases earnings by about 78 (FWLS) or 75 (FNSEM) percent, compared to holding no qualification. Vocational training alone increases earnings by about 16 (FWLS) or 22 (FNSEM) percent.

In the last column, we use the more detailed educational information in the FNSEM, and decompose educational attainments into overseas and UK qualifications. We find that the coefficients on UK qualifications are larger than overseas ones, and the differences are significant for degree and Alevels. An overseas degree still increases earnings by 44%, compared to those with no qualification; however, returns to a degree obtained in the UK are about 73% higher. A slightly lower difference exists for respondents with low Alevels (about 27%).

The coefficients on the ethnicity dummies indicate large wage differences between ethnic groups. Like in the language and employment equations, Bangladeshis are the most disadvantaged group. Conditional on education, age and years of residence, wages are 54 percent lower than those of the most successful group, the Chinese

(FNSEM). In both data sets, Indians and Caribbeans are receiving wages which are more than 30% higher than Bangladeshi wages; also, Pakistani wages are about 20% higher.

Table 6: Earnings Regressions

Variable	FWLS						FNSEM			
	1		2		3		Speaking			
							All Qualifications		UK/nonUK Q	
	Coeff	t-ratio	Coeff	t-ratio	Coeff	t-ratio	Coeff	t-ratio	Coeff	t-ratio
cons	3.551	8.622	3.577	8.683	3.546	8.586	3.330	12.786	3.246	12.335
sex	0.238	3.298	0.251	3.526	0.238	3.295	0.147	3.360	0.155	3.568
married	-0.010	-0.121	-0.008	-0.097	-0.008	-0.095	0.161	2.886	0.146	2.620
degree	0.786	7.505	0.788	7.428	0.781	7.350	0.702	13.146	–	–
Alevtea	0.206	2.279	0.202	2.175	0.201	2.166	0.433	7.674	–	–
OlevCSE	0.169	1.856	0.172	1.878	0.166	1.809	0.228	4.766	–	–
Edegree	–	–	–	–	–	–	–	–	0.729	11.786
Elevtea	–	–	–	–	–	–	–	–	0.418	7.031
ElevCSE	–	–	–	–	–	–	–	–	0.233	4.229
Fdegree	–	–	–	–	–	–	–	–	0.437	6.093
Flevtea	–	–	–	–	–	–	–	–	0.155	1.809
FlevCSE	–	–	–	–	–	–	–	–	0.136	2.488
age	0.038	1.639	0.036	1.524	0.038	1.614	0.029	2.141	0.036	2.565
agesq	-0.045	-1.550	-0.042	-1.441	-0.044	-1.522	-0.035	-2.108	-0.040	-2.410
yearst	0.026	1.749	0.030	2.085	0.027	1.767	0.038	4.607	0.033	3.913
yearst ²	-0.050	-1.403	-0.035	-1.640	-0.050	-1.419	-0.061	-2.836	-0.056	-2.613
black	0.302	2.285	0.327	2.504	0.301	2.271	0.479	5.840	0.480	5.817
afroas	0.081	0.647	0.109	0.882	0.083	0.660	0.358	4.972	0.369	5.115
indian	0.311	2.735	0.329	2.928	0.310	2.728	0.310	4.215	0.359	4.937
pakista	0.239	2.015	0.251	2.126	0.239	2.012	0.186	2.399	0.222	2.889
Chinese	–	–	–	–	–	–	0.562	6.232	0.527	5.787
speak	0.204	1.774	–	–	0.171	1.062	0.162	2.706	0.170	2.850
write	–	–	0.149	1.444	0.040	0.282	–	–	–	–
No. of Obs.	254		254		254		964		964	
	Excluded categories: no education qualification, Bangladeshi.									

In both data sets, we find large and significant coefficients on the English fluency variables. The point estimates in the FNSEM and FWLS are quite similar, and

indicate that English language proficiency is associated with 16 percent (FNSEM) or 20 percent (FWLS) higher wages. Interestingly, and different from the employment equation, fluency seems to be more important for wages than literacy, as is indicated by the results both in columns 2 and 3.

5.3 Earnings Assimilation and Language

Both the FNSEM and the FWLS contain comparison samples on native born non-white minorities, and on white natives. Not much is known about the performance of ethnic minority immigrants, relative to native born minorities, and to white natives. Earlier work by Blackaby et al. (1994) suggests that black minorities in the UK are disadvantaged, compared to the white majority population. Bell (1997) adds to this evidence by breaking up the black minority into foreign-born and native-born individuals. He shows that the foreign-born have a substantial wage disadvantage, compared to black natives.

Using the samples on white and ethnic minority natives in the FNSEM and the FWLS, we use simple regression analysis to compare earnings paths of minority immigrants to native born minorities, and to white native born individuals.

Results of basic specifications are reported for all three groups in tables A4 and A5. For the FNSEM, we run regressions including, and excluding minority dummies. Results in column 3 in table A4 are identical to results in Table 6.

As for minority immigrants, there are differences between ethnic groups also in the native born sample. The coefficients on the ethnicity dummies are not estimated with great precision, due to the small sample size, but the estimates indicate a similar pattern to the immigrant sample. Again, the Chinese seem to be most successful; furthermore, the Bangladeshis seem to be the most disadvantaged group. We have also included ethnicity dummies in the FWLS native minority sample (results not reported). Since the sample is very small, coefficients are not significant, but they are all positive (again, we exclude the Bangladeshi group).

To illustrate wage paths and wage assimilation of immigrants to the native born, we plot log weekly earnings against age in the labour market, where we assume for immigrants that the age at entry is 20 years (which is close to the sample average for both samples). Graphical illustrations are based on results in columns 1, 2, and 4, of table A4, and on results in table A5, where no origin dummies are included. We compute entry wages for each group as the sum of the regression coefficients, weighted by the respective sample means, and setting the language dummy equal to zero, or to one for immigrants. Accordingly, wage paths for immigrants represent an average of the paths of individuals of different ethnic groups (according to the sample data), who do (not) speak the English language well. Figures (1) and (2) display the age-earnings profiles.

Log wage profiles for white natives and minority natives are remarkably similar

Figure 1: Log Weekly Earnings - Age Profiles, FNSEM

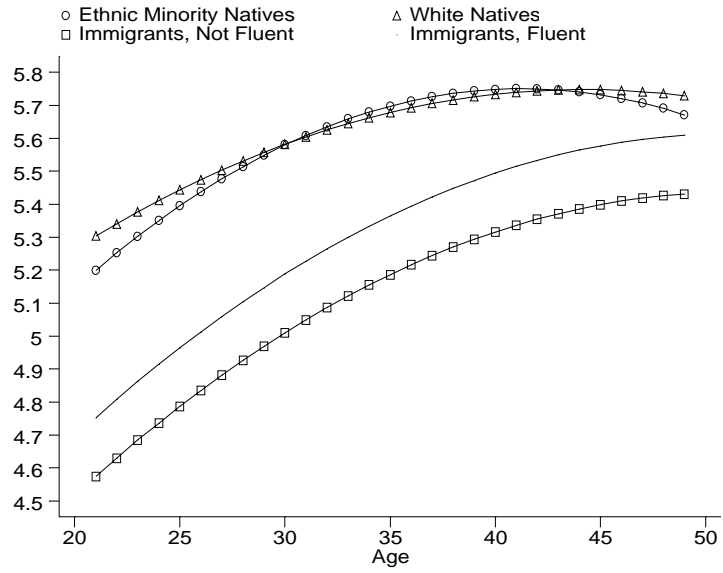
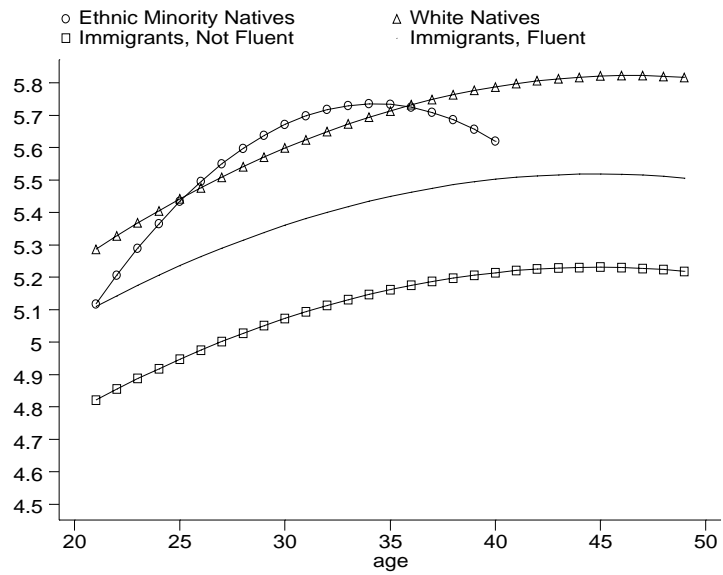


Figure 2: Log Weekly Earnings - Age Profiles, FWLS



for both samples, with white native profiles being almost identical.⁶ Entry level wages for minorities are slightly lower than those for whites.

For the FNSEM, the entry level wage difference for minority immigrants who are not fluent in English is quite dramatic. The difference in log entry wages is about 0.74 between immigrants and white natives, and 0.62 between immigrants and native minorities. The gap is closing, but it never disappears. At the average age in our sample (which is 42 years), the log difference is still 0.39 (which corresponds to a 46 percent wage gap). If we compute the wage gap between specific ethnic groups and white natives at the average sample age, we obtain a more diverse picture. For instance, while the log wage gap between a (non-fluent) Bangladeshi immigrant and a native is about 0.7, it is only 0.14 between a native white and a Chinese immigrant, or 0.22 between a native white and an Indian immigrant. Ethnic origin plays a major role in predicting disadvantages of immigrant minorities.

For immigrants who are fluent in English, the wage disadvantage reduces quite considerably. For some ethnic groups, it disappears completely. For Chinese immigrants who are fluent in English, the wage gap is positive, and it nearly disappears for Indians.

For the FWLS, the initial log earnings gap between white natives, and non-fluent minority immigrants is about 0.46, which is slightly lower, but still considerable.

⁶For the FWLS, we plot the minority natives only up to age 40. After that age there is hardly any data support.

The gap seems to widen slightly over the immigrant's migration history. The larger earning gap in the FNSEM may be related to the sample composition. Remember that computations are for the average for the different ethnic groups, according to the ethnicity composition in our data. Since Bangladeshis (the most disadvantaged group) are oversampled in the FNSEM, the estimates may indeed slightly overstate the wage disadvantage of the average ethnic minority immigrant. Also, individuals in the FNSEM are on average slightly older than those in the FWLS. Still, the results suggest that minority immigrants are, on average, heavily disadvantaged, not only at entry to the labour market, but over their entire migration history. On the other side, minority native born individuals do surprisingly well. This underlines the importance of distinguishing between native and foreign born minorities when analysing their economic assimilation.

As for the FNSEM, the wage gap closes quite considerably for immigrants who are fluent in English. For the FWLS, this reduction amounts to half of the initial wage gap.

Based on these estimates, some simple back of the envelope calculations, based on FNSEM results, suggest that an immigrant, who arrives at age 20, and who does not acquire fluency in English over the first ten years of his/her migration history, loses about £10,000, compared to an immigrant who arrives with language fluency. This is a very considerable amount, given that the earnings accumulated

by an average immigrant over that period is £68,000. Similar calculations for the FWLS suggest an earnings loss of even £21,000, with the average earnings over the 10 years period being equal to £86,000.

5.4 Measurement error and unobserved ability

The results suggest that language is a most important determinant of labour market success, and that the earnings disadvantage of minority immigrants relative to native born individuals is largely reduced if the individual is proficient in the English language. The estimated coefficients in outcome equations of the sort estimated above may be seriously biased however due to unobserved heterogeneity and measurement error. In this section, we address this problem.

To address the bias due to measurement error, we use the language in which the interview has been performed as an instrument for the assessed language proficiency of the respondent. This information is available in the FNSEM. In all areas with a minority density above 0.5% (which includes 97% of the sample individuals), there was an initial screening interview with the interviewee. In the case of poor fluency, the interviewers were chosen to be fluent in the language of the respondents. During the interview, interviewers decided about the extent to which English could be used in the interview, and we have information as to whether the interview was held wholly in English, partly in English, or wholly in the individual's mother tongue.

We use this information as instrument for the assessed language fluency of the individual.⁷

This variable does not eliminate the bias due to unobserved heterogeneity, however. Our estimation strategy for the unobserved heterogeneity component is a type of matching estimator, as explained above. We add partner and family variables to the main equation. If these variables explain some of the variation in the unobserved error component which is correlated with language proficiency, this estimator reduces (or eliminates) the bias due to unobserved heterogeneity.

In Table 7, we display results for the employment and earnings equations. We report marginal effects, evaluated at sample means. The first column corresponds to the simple estimates in Tables 5 and 6. Column 2 reports IV estimates, where we use interview language as an instrument. This eliminates the bias due to measurement error, but not the bias due to unobserved heterogeneity. We first discuss results for the employment equation. We estimate the employment equation as a bivariate probit model, using maximum likelihood. Marginal effects increase from 17 percent to 22 percent, and the correlation coefficient is negative and significant, indicating a downward bias due to measurement error.

Adding partner variables reduces our sample to married couples, and we report

⁷Since fluency in the FNSEM is also interviewer assessed, measurement error in the two language variables may be correlated. Should this be the case, our estimation strategy does only reduce, but not entirely eliminate the bias. Therefore, the estimates we obtain are a lower bound.

in column 3 the marginal effects for this restricted sample. For completeness, we also report the corresponding coefficient for the FWLS sample. Both are lower than the coefficient in the overall sample. Column 4 reports results without conditioning on family background, but using the information on interview language as instruments to eliminate the measurement error bias. The coefficient increases by about 4 percentage points, as compared to results in column 3. Finally, column 5 reports results when we include family background characteristics, and use IV estimation. Inclusion of the family variables reduces the coefficient by 2 percentage points, but the estimate is still larger than that of the simple regression model. The correlation coefficient is negative in both regressions.

Table 9: Language and Employment Probabilities

Variable	All				Married Sample						
	1		2		3		4		5		
Language and Employment Probabilities											
	M.E.	t-ratio	M.E.	t-ratio	M.E.	t-ratio	M.E.	t-ratio	M.E.	t-ratio	
FNSEM											
Speak	0.172	5.99	0.218	6.27	0.136	4.74	0.171	4.64	0.153	3.93	
ρ			-0.425	-2.35			-0.330	-1.78	-0.320	-1.66	
N. Obs.	2092		2092		1653		1653		1653		
FWLS											
Speak	0.222	4.24			0.186	3.34					
N. Obs.	849				643						
Language and Earnings											
	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio	
FNSEM											
Speak	0.162	2.706	0.223	2.55	0.209	3.175	0.271	2.857	0.255	2.64	
ρ			-0.030	0.12			-0.084	0.54	-0.06	0.41	
N. Obs.	964		964		777		777		777		
FWLS											
Speak	0.204	1.774			0.229	1.968					
N. Obs.	259				257						
1: Grouped Regression. 2: Grouped Regression/Probit, using interview language as instruments. 3: Grouped Regression. 4: Grouped Regression/Probit, using interview language as instruments. 5: Grouped Regression/Probit, including partner variables, using interview language as instruments.											

In lower panel of table 9, we report the corresponding results for the earnings equation. We estimate the model using maximum likelihood, using a grouped regression/probit mixture. Instrumenting language increases the coefficient by about 6 percentage points, indicating a strong downward bias due to measurement error in straightforward estimations. Other than for the employment equation, the language coefficient increases slightly when considering married individuals only (column 3). Column 4 shows that instrumenting language for measurement error leads, as before, to a substantial increase in the coefficient for this subsample (by about 7 percentage

points). In column 5 we add partner variables to account for unobserved heterogeneity. As in the employment estimation, the coefficient reduces slightly in size, but is still larger than the coefficient in straightforward regressions.

These results suggest that measurement error leads to quite a substantial downward bias in the effect of language on labour market outcomes. They indicate that language is more important than straightforward regressions suggest, and that the simple calculations we have performed in the previous section on the earnings loss and employment probabilities due to a lack of language fluency may be underestimates, rather than overestimates.

6 Discussion and Conclusion

Based on two recent UK surveys, which provide sufficiently large samples to study ethnic minority immigrants, we analyse the determinants of English language fluency for ethnic minority migrants in the UK, the effect of language on labour market outcomes, and the contribution of language to reduce the earnings gap between immigrants, and native born individuals.

We find that in simple regressions, language proficiency is strongly associated with higher employment probabilities, and with higher earnings. When we compare earnings of minority immigrants with those of white and minority natives, our data indicates a substantial disadvantage for the average immigrant, compared to white

natives. The same patterns of ethnic differences in earnings can be observed between ethnic minority immigrants, and ethnic minority natives. The earnings disadvantage does not disappear over time. There are large earnings differences between individuals who belong to different minority communities - Chinese immigrants, for instance, have about 50 percent higher earnings than Bangladeshi immigrants. The earnings gap is therefore much smaller (and even reverses) for some minority immigrants, while it is larger for others. Language proficiency is an important factor in reducing these earnings differences.

Language effects may be under- or overestimated in straightforward regressions, due to unobserved heterogeneity and measurement error. We address both these issues. To reduce the bias due to unobserved heterogeneity components, we use a simple matching approach. To address the bias due to measurement error, we use information about the interview language as an instrument of the individual's language potential. Our results indicate that measurement error leads to a downward bias in language estimates for both outcomes. Straightforward regressions seem to underestimate this effect by about 6 percentage points.

Table 10: Language and Earnings, different countries

	UK FNSEM	UK FWLS	Germany ¹	United States ²	Canada ²	Australia ²	Israel ²
OLS	0.209	0.229	0.040	0.169	0.122	0.083	0.11
StdE	(0.066)	(0.116)	(0.011)	(0.013)	(0.050)	(0.017)	(0.009)
IV (M.E.)	–	0.271	0.129	–	–	–	–
StdE	–	(0.090)	(0.017)	–	–	–	–
IV, Matching*	–	0.255	0.097	–	–	–	–
StdE	–	(0.112)	(0.064)	–	–	–	–

Standard errors in brackets

*: Married Subsample

¹: Dustmann and van Soest (1999). ²: Chiswick and Miller (1995).

How do results on the effects of language on earnings compare to those of other countries? In table 10, we display estimates obtained for Germany, the US, Canada, Australia, and Israel. The numbers reveal that the association between language and earnings is largest for the UK, followed by the US and Canada. Much smaller effects are found for Germany, Australia, and Israel. Accordingly, language is most important in the UK labour market. A reason for the relatively larger estimates, compared to Germany and Israel, may be that English is a more dominant language than German or Hebrew; thus, coming from a minority population, it may be possible to communicate with other minorities, or majority individuals, in a third language (which will be English in most cases). However, a bad fluency in English, if English is the majority language, is likely to have more serious consequences. Not only does it hinder communication with majority individuals, but also with other minorities (or even with minority individuals of the same ethnic origin who speak a different language). The relatively large coefficients found for the US are in line with this interpretation.

There is only one study (for Germany) which instruments for measurement error in the language variables. Instrumenting leads to quite a dramatic increase in the coefficient, but estimates are still much lower than those for the UK. In both studies, coefficient estimates reduce slightly in size when matching on partner variables.

We may conclude that language proficiency is an important factor for economic success of immigrants in the UK, and the earnings losses immigrants experience due to a lack of language fluency are substantial. Language seems to be more significant in affecting earnings of UK immigrants than those of immigrants to other countries. As a consequence, it may be worthwhile to implement schooling centers which support immigrants in their acquisition of the English language at an early stage of their migration history. Given the substantial earnings disadvantages immigrants experience due to a lack of English fluency, there is clearly an incentive for the migrants to bear part of the costs of language education.

7 References

- Bell, B.D. (1997): "The Performance of Immigrants in the United Kingdom: Evidence from the GHS", *The Economic Journal*, 107, pp. 333-344.
- Blackaby, D.H., Clark, K., Leslie, D.G., Murphy, P.D. (1994): "Black-white male earnings and employment prospects in the 1970s and 1980s Evidence for Britain", *Economics Letters*, 46, pp. 273-279.
- Blackaby, D.H., Drinkwater, S., Leslie, D.G., Murphy, P.D. (1997): "A Picture of Male and Female Unemployment Among Britain's Ethnic Minorities", *Scottish Journal of Political Economy*, 44(2), pp. 182-197.
- Borjas (1985), Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants, *Journal of Labour Economics*, 3, 463-489.
- BORJAS, G.J. (1987): "Self-Selection and the Earnings of Immigrants," *American Economic Review*, 77, 531-553
- Borjas, G.J. (1994): "The Economics of Immigration", *Journal of Economic Literature*, 32, pp. 1667-1717.
- CHISWICK, B.R. (1978): "The Effect of Americanization on the Earnings of Foreign-Born Men," *Journal of Political Economy*, 86, 897 - 921.

- Chiswick, B.R. (1980): "The Earnings of White and Coloured Male Immigrants in Britain", *Economica*, 47, pp. 81-87.
- Chiswick, B.R. (1991): "Speaking, Reading, and Earnings among Low-skilled Immigrants", *Journal of Labor Economics*, 9, pp. 149-170.
- Chiswick, B.R. (1994): "Language Choice Among Immigrants in a Multi-Lingual Destination", *Journal of Population Economics*, 7, pp. 119-131.
- Chiswick, B.R., Miller, P.W. (1995): "The Endogeneity between Language and Earnings: International Analyses", *Journal of Labor Economics*, 13, pp. 246-288.
- Chiswick, B.R., Cohen, Y., Zach, T. (1997): "The Labor Market Status of Immigrants: Effects of the Unemployment Rate at Arrival and Duration of Residence", *Industrial and Labor Relations Review*, 50(2), pp. 289-303.
- Dearden, L. (1999): "Qualifications and Earnings in Britain: How reliable are Conventional OLS estimates of the Returns to Education?", IFS working paper No. W99/7.
- Dustmann, C. (1994): "Speaking Fluency, writing fluency and earnings of migrants", *Journal of Population Economics*, 7, pp. 133-156.
- Dustmann, C. van Soest, A. (1998): "Language and the Earnings of Immigrants", Centre for Economic Policy Research Discussion Paper No. 2012.

- Dustmann, C., Fabbri, F., and Preston, I. (2000): "The Causal Effect of Ethnic Concentration on Language", mimeo, University College London.
- Eckstein, Z. and Y. Weiss (1999): "The Integration of Immigrants from the Former Soviet Union in the Israeli Labour Market", mimeo, Tel Aviv University.
- Green, A.E., Owen, D.W. (1995): "Ethnic Minority Groups in Regional and Local Labour Markets in Britain: A Review of Data Sources and Associated Issues", *Regional Studies*, 29(8), pp. 729-735.
- Hatton, T.J., and S. Wheatley Price (1999), "Migration, Migrants and Policy in the United Kingdom", IZA discussion paper No. 81.
- Lazear, E.P. (1995): "Culture and Language", NBER working paper No. 5249.
- Modood, T., Berthoud, R. (1997): *Ethnic Minorities in Britain*, Policy Studies Institute, London.
- Nickell, S.J. (1980): "A Picture of Male Unemployment In Britain", *Economic Journal*, 90, pp. 776-794.
- Rivera-Batiz, F. L. (1990): "English language proficiency and the economic progress of immigrants", *Economics Letters*, 34, pp. 295-300.

- Rosenbaum, P., Rubin, D.B. (1983): "The central role of the propensity score in observational studies for causal effects", *Biometrika*, 70, pp. 41-55.
- Shields, M.A., Wheatley Price, S. (1998): "The Earnings of Male Immigrants in England: Evidence From the Quarterly LFS", *Applied Economics*, forthcoming.
- Shields, M.A., Wheatley Price, S. (1999): " The English Language Fluency and Occupational Success of Ethnic Minority Immigrant Men Living in English Metropolitan Areas", University of Leicester Discussion Paper.
- Sly, F. (1995):"Ethnic groups and the labour market", *Employment Gazette*, May, pp. 147-158.
- Smith, P. and G. Prior (1996), "The Fourth National Survey of Ethnic Minorities: Technical Report", Social and Community Planning Research.
- Stewart, M.B. (1983): "On Least Squares Estimation when the Dependent Variable is Grouped", *Review of Economic Studies* 50:737-753.
- Wheatley Price, S. (1997): "The Employment Adjustment of Male Immigrants to the English Labour Market", mimeo, University of Sheffield.

Appendix: Tables

Table A1: Language Information

	All groups	Caribbean	Indian	Afroasian	Pakistani	Bangladeshi	Chinese
	English as 1st language, FWLS						
Yes	26.92	91.41	27.46	28.33	10.89	5.74	-
No	73.08	8.59	72.54	71.67	89.11	94.26	-
	Speaking, FWLS						
Very well	37.81	54.55	50.44	64.77	38.16	25.93	-
Quite well	23.12	13.64	27.43	27.27	26.05	18.46	-
Not well	20.12	18.18	18.14	5.68	21.32	22.82	-
Hardly	11.69	13.64	3.54	2.27	10	18.46	-
Not at all	7.26	-	0.44	-	4.47	14.32	-
	Reading, FWLS						
Very well	34.64	40.91	48.67	61.36	33.16	24.07	-
Quite well	21.12	18.18	23.89	26.14	21.58	18.67	-
Not well	15.86	22.73	14.16	7.95	17.11	16.8	-
Hardly	13.19	9.09	7.96	1.14	14.47	17.01	-
Not at all	15.19	9.09	5.31	3.41	13.68	23.44	-
	Writing, FWLS						
Very well	32.39	40.91	45.13	56.82	29.47	23.86	-
Quite well	19.2	18.18	21.68	23.86	20.79	15.98	-
Not well	16.61	22.73	15.49	13.64	18.16	16.18	-
Hardly	12.77	4.55	11.06	2.27	13.68	15.15	-
Not at all	19.03	13.64	6.64	3.41	17.89	28.84	-
	Speaking, FNSEM						
Fluent	48.73	89.65	39.98	65.63	25.56	25.97	56.59
Fairly	20.4	9.62	24.37	19.2	25.56	23.02	12.64
Slightly	21.2	-	25.84	11.76	32	34.25	18.13
Not at all	9.67	-	9.81	3.41	16.88	16.76	12.64

Table A2: Age and Time Patterns

Ethnicity	Age Migration		Year		Age	
	FWLS	FNSEM	FWLS	FNSEM	FWLS	FNSEM
Carib	19.460	20.379	1964.2	63.435	49.755	50.927
	8.840	10.186	7.968	7.772	12.199	13.933
Indian	18.971	23.892	1973.2	72.431	40.299	45.145
	10.235	13.535	9.750	9.515	11.004	14.384
AfroAsia	17.813	20.662	1974.1	73.828	37.976	40.735
	10.320	12.702	7.035	7.221	10.393	13.026
Pakista	18.167	20.424	1976.4	74.207	35.870	39.672
	9.374	11.275	9.634	9.677	11.953	13.738
Bangla	18.676	20.579	1979.6	77.695	33.266	36.645
	9.374	10.545	9.046	9.637	13.985	14.156
Chinese	-	22.088	-	76.35	-	39.641
	-	11.860	-	8.670	-	12.532
All	18.663	21.418	1974.7	72.414	38.308	42.707
	10.084	11.918	1.402	9.944	13.587	14.572

Table A3: Attrition

Variable	FWLS					FNSEM				
	Missing		Report Earnings		Diff.	Missing		Report Earnings		Diff.
	Mean	StdD	Mean	StdD	t-value	Mean	StdD	Mean	StdD	t-value
sex	0.730	0.444	0.618	0.486	2.584	0.675	0.468	0.695	0.460	-0.460
married	0.802	0.398	0.767	0.423	0.914	0.870	0.336	0.837	0.368	1.006
nchild	1.995	1.712	1.480	1.341	3.539	1.511	1.592	1.507	1.466	0.027
degree	0.110	0.314	0.153	0.361	-1.368	0.226	0.419	0.195	0.396	0.810
Alevtea	0.144	0.352	0.212	0.409	-1.920	0.164	0.371	0.157	0.364	0.203
OlevCSE	0.278	0.449	0.208	0.407	1.738	0.249	0.433	0.262	0.440	-0.318
age	38.274	10.456	37.704	10.006	0.594	41.556	10.378	39.214	10.253	2.426
yearstay	22.783	9.453	22.303	9.3220	0.546	23.349	8.246	21.780	9.045	1.947
black	0.182	0.387	0.204	0.404	-0.595	0.109	0.312	0.217	0.412	-3.203
afroas	0.129	0.336	0.145	0.353	-0.497	0.219	0.414	0.242	0.428	-0.585
indian	0.264	0.442	0.338	0.474	-1.732	0.408	0.492	0.227	0.419	4.202
pakista	0.278	0.449	0.173	0.379	2.680	0.197	0.398	0.147	0.354	1.411
Chinese	-	-	-	-		0.024	0.156	0.073	0.260	
speak	0.865	0.342	0.877	0.327	-0.382	0.879	0.325	0.872	0.333	0.227
write	0.793	0.405	0.850	0.357	-1.586	-	-	-	-	
No. of Obs.	208		254			401		1012		
	Note: t-statistics computed as $(m_1 - m_2)/\sqrt{se_1^2 + se_2^2}$, where m_i, se_i are means and standard errors of the two sample values, respectively.									

Table A4: Immigrants, Whites, Native born Minorities, FNSEM

Variable	White Natives		Immigrants				Native Born Minorities			
	Coeff	t-ratio	Coeff	t-ratio	Coeff	t-ratio	Coeff	t-ratio	Coeff	t-ratio
cons	3.616	20.068	3.348	12.384	3.330	12.786	3.165	7.389	3.706	7.724
sex	0.287	9.080	0.056	1.315	0.147	3.360	0.090	1.701	0.090	1.740
age	0.073	7.667	0.046	3.257	0.029	2.141	0.110	3.922	0.062	2.128
agesq/100	-0.082	-6.916	-0.052	-3.078	-0.035	-2.108	-0.133	-2.895	-0.000	-1.582
degree	0.705	14.015	0.762	13.963	0.702	13.146	0.577	6.283	0.583	6.473
AlevTea	0.433	9.549	0.476	8.193	0.433	7.674	0.318	3.680	0.244	2.876
voc	0.189	4.621	0.258	5.198	0.228	4.766	0.186	2.550	0.112	1.546
married	0.097	2.872	0.137	2.435	0.161	2.886	0.036	0.663	0.100	1.813
yearstay	–	–	0.033	3.941	0.038	4.607	–	–	–	–
yearst2/100	–	–	-0.041	-1.928	-0.061	-2.836	–	–	–	–
indian	–	–	–	–	0.479	5.840	–	–	0.191	0.818
afroas	–	–	–	–	0.358	4.972	–	–	0.312	1.200
pakista	–	–	–	–	0.310	4.215	–	–	0.084	0.353
carib	–	–	–	–	0.186	2.399	–	–	0.443	1.886
chinese	–	–	–	–	0.562	6.232	–	–	0.435	1.570
speak	–	–	0.178	2.903	0.162	2.706	–	–	–	–
No. of Obs.	921		964		964		322		322	

Excluded categories: no education qualification, Bangladeshi (cols 3,5).

Table A5: Immigrants, Whites, Native born Minorities, FWLS

Variable	White Natives		Immigrants		Native Born Minorities	
	Coeff	t-ratio	Coeff	t-ratio	Coeff	t-ratio
cons	3.499	28.246	3.670	9.104	1.346	1.771
sex	0.250	11.505	0.208	3.066	0.242	2.680
age	0.076	11.673	0.787	7.776	0.240	4.724
agesq/100	-0.082	-10.071	0.243	2.726	-0.351	-4.121
degree	0.719	18.899	0.152	1.707	0.439	2.354
AlevTea	0.437	13.103	0.055	2.475	0.160	1.170
voc	0.278	8.556	-0.065	-2.291	0.176	1.278
married	0.073	3.138	-0.094	-1.167	-0.042	-0.402
yearstay	–	–	0.006	0.454	–	–
yearst2/100	–	–	-0.007	-0.215	–	–
speak	–	–	0.288	2.697	–	–
No. of Obs.	2559		255		142	

Excluded categories: no education qualification.