research paper series

Globalisation and Labour Markets Programme

Research Paper 2001/15

The English Language Fluency and Earnings of Ethnic Minorities in Britain

By J. Lindley

Leverhulme Centre
for Research on Globalisation and Economic Policy

The Centre acknowledges financial support from The Leverhulme Trust
under Programme Grant F114/BF

The Authors

Joanne Lindley is a Research Fellow in The Leverhulme Centre for Research on Globalisation and Economic Policy, University of Nottingham.

Acknowledgements

The author thanks Derek Leslie and Ken Clark for advice and comments. *The Fourth National Survey of Ethnic Minorities* was made available through the ESRC Data Archive.

The English Language Fluency and Earnings of Ethnic Minorities in Britain

by

J. Lindley

Abstract

This study addresses two issues. First it estimates how much of the male and female ethnic earnings gap is the result of an advantage in the English language and whether there is an earnings penalty to non-whites, over and above this. Lack of fluency is shown to have a highly significant impact on the earnings of ethnic minorities in Britain, although the language penalty is much greater for women than it is for men. Moreover, only foreign born non-white males exhibit lower earnings once language fluency is taken into consideration, whilst British born females exhibit higher earnings. So the evidence here suggests that non-white earnings are assimilating towards those of whites and that lower female non-white earnings are a direct result of a lack of fluency rather than ethnicity. Secondly, the study will try to measure any endogenous bias associated with the non-fluency earnings penalty. Controlling for the endogeneity between language fluency and earnings is shown to be problematic. Estimates suggest that single equation earnings functions underestimate the true language fluency penalty for males, and overestimate the fluency penalty for females. Finally education and fluency are not surprisingly shown to be complementary.

Outline

Non-Technical Summary

Ethnic assimilation is the convergence of immigrant human capital and economic outcomes, towards those of the indigenous native population. This becomes increasingly more important with European Economic convergence and the growth of labour movements. This study addresses one important part of ethnic assimilation, that of language fluency assimilation. The main aim is to estimate the determinants of English language fluency and to assess the impact of poor language ability on earnings in Britain. The concern is with both British and foreign born non-whites, since this group contains both first and second generation immigrants. In order to detect any ethnic disadvantage over and above human capital differences (including lack of language fluency), outcomes for non-whites are compared to those for whites.

The Fourth National Survey of Ethnic Minorities was conducted in 1994 by the Policy Studies Institute to investigate the social and economic conditions of the ethnic minorities of England and Wales. Electoral wards that contained a high percentage of ethnic minorities were over-sampled. A key advantage of the FNSEM was that interviews were conducted both in English and, if necessary, the main language of the respondent. This enabled the FNSEM to capture those from the ethnic minority population who had poor English language fluency. As a consequence, the FNSEM includes information on English language speaking ability.

The study adopts a human capital approach to assess the impact of non-fluency on the earnings of ethnic minority males and females. All males are shown to have a higher propensity for non-fluency relative to Black Caribbean males, with Pakistanis and Indians displaying the highest propensity for non-fluency. All females have a higher propensity to be non-fluent relative to Black Caribbean s. Indians and Chinese display the highest non-fluency, with African Asian women displaying the lowest. Compared to males, the ethnic penalties for female non-fluency appear to be generally bigger. Furthermore, it is shown that high ethnic density implies lower language fluency. This supports the idea that immigrants in areas of high ethnic concentration should be less proficient in the host language simply because there is less necessity to acquire language skills.

Lack of fluency is shown to have a highly significant impact on the earnings of ethnic minorities in Britain, and this is especially true for female non-whites. After qualifications, it is the language penalty that has the biggest impact on female earnings and local unemployment rates that have the biggest affect on the earnings of men. However, there is some evidence that non-white earnings are assimilating towards those of whites and that lower female non-white earnings are a direct result of a lack of fluency rather than ethnicity. Finally education and fluency are not surprisingly shown to have complementary effects on earnings.

The Authors

Joanne Lindley is a Research Fellow in The Leverhulme Centre for Research on Globalisation and Economic Policy, University of Nottingham.

Acknowledgements

The author thanks Derek Leslie and Ken Clark for advice and comments. *The Fourth National Survey of Ethnic Minorities* was made available through the ESRC Data Archive.

The English Language Fluency and Earnings of Ethnic Minorities in Britain

by

J. Lindley

Abstract

This study addresses two issues. First it estimates how much of the male and female ethnic earnings gap is the result of an advantage in the English language and whether there is an earnings penalty to non-whites, over and above this. Lack of fluency is shown to have a highly significant impact on the earnings of ethnic minorities in Britain, although the language penalty is much greater for women than it is for men. Moreover, only foreign born non-white males exhibit lower earnings once language fluency is taken into consideration, whilst British born females exhibit higher earnings. So the evidence here suggests that non-white earnings are assimilating towards those of whites and that lower female non-white earnings are a direct result of a lack of fluency rather than ethnicity. Secondly, the study will try to measure any endogenous bias associated with the non-fluency earnings penalty. Controlling for the endogeneity between language fluency and earnings is shown to be problematic. Estimates suggest that single equation earnings functions underestimate the true language fluency penalty for males, and overestimate the fluency penalty for females. Finally education and fluency are not surprisingly shown to be complementary.

Outline

1. Introduction
2. The Theory of Earnings and Language Fluency
3. The Data
4. Earnings Functions for Whites and Non-whites
5. The Simultaneous Equation Approach
6. The Selection Equation Approach
7. Concluding Comments

Non-Technical Summary

Ethnic assimilation is the convergence of immigrant human capital and economic outcomes, towards those of the indigenous native population. This becomes increasingly more important with European Economic convergence and the growth of labour movements. This study addresses one important part of ethnic assimilation, that of language fluency assimilation. The main aim is to estimate the determinants of English language fluency and to assess the impact of poor language ability on earnings in Britain. The concern is with both British and foreign born non-whites, since this group contains both first and second generation immigrants. In order to detect any ethnic disadvantage over and above human capital differences (including lack of language fluency), outcomes for non-whites are compared to those for whites.

The Fourth National Survey of Ethnic Minorities was conducted in 1994 by the Policy Studies Institute to investigate the social and economic conditions of the ethnic minorities of England and Wales. Electoral wards that contained a high percentage of ethnic minorities were over-sampled. A key advantage of the FNSEM was that interviews were conducted both in English and, if necessary, the main language of the respondent. This enabled the FNSEM to capture those from the ethnic minority population who had poor English language fluency. As a consequence, the FNSEM includes information on English language speaking ability.

The study adopts a human capital approach to assess the impact of non-fluency on the earnings of ethnic minority males and females. All males are shown to have a higher propensity for non-fluency relative to Black Caribbean males, with Pakistanis and Indians displaying the highest propensity for non-fluency. All females have a higher propensity to be non-fluent relative to Black Caribbean s. Indians and Chinese display the highest non-fluency, with African Asian women displaying the lowest. Compared to males, the ethnic penalties for female non-fluency appear to be generally bigger. Furthermore, it is shown that high ethnic density implies lower language fluency. This supports the idea that immigrants in areas of high ethnic concentration should be less proficient in the host language simply because there is less necessity to acquire language skills.

Lack of fluency is shown to have a highly significant impact on the earnings of ethnic minorities in Britain, and this is especially true for female non-whites. After qualifications, it is the language penalty that has the biggest impact on female earnings and local unemployment rates that have the biggest affect on the earnings of men. However, there is some evidence that non-white earnings are assimilating towards those of whites and that lower female non-white earnings are a direct result of a lack of fluency rather than ethnicity. Finally education and fluency are not surprisingly shown to have complementary effects on earnings.

1. Introduction

Following, the pioneering work of Chiswick (1978), the impact of language fluency on economic outcomes has been explored fairly extensively. Blackaby *et al.* (2001) explored the impact of fluency on the male ethnic wage gap in Britain. They find only a small language penalty for males, which is significantly less than the ethnic penalty and any other characteristic effect. It has been argued however, that individuals have an economic incentive to acquire language skills. Following Chiswick and Miller (1995), Shields and Price (2001) tried to correct for this endogeneity by using instrumental variables. They find that fluency is the second most important determinant of occupational success amongst British immigrant men. This paper updates earlier work by estimating separate earnings functions for males and females, including both British born and immigrant non-whites, and addressing the issue of language endogeneity. First, single equation earnings functions are estimated for both whites and non-whites. Secondly, a simultaneous equation model is estimated for non-whites only. This involves a two stage least squares approach using an earnings equation and a probit equation for English language proficiency. Finally a language fluency selection model is estimated.

The paper proceeds as follows. Section 2 provides an insight into the theory of language and develops the estimation equations. Section 3 describes the data source and presents some descriptive statistics. Section 4 analyses how language proficiency affects the earnings of non-whites relative to whites, using single equation estimates. Sections 5 and 6 address the issue of endogeneity between language and earnings. Section 5 adopts a simultaneous estimation approach, whilst section 6 reports the selection equation estimates. The final section summarises the results obtained.

2. The Theory of Earnings and Language Fluency

The factors that determine language acquisition have been documented by Chiswick and Miller (1995). These are collated into three categories. The first set of factors increase immigrant efficiency in learning to speak the host language. These include ability, age, colonial ties of the native country with the host country and the linguistic distance between native tongue and host language. The second set of factors correspond to the immigrants exposure to the host language before and after migration. These include pre-migration and post-migration qualifications, whether the host language was frequently used in the country of origin, time spent in the destination country and whether the immigrant lives in an area that is highly concentrated in terms of members of the same ethnic group. Finally there are economic incentives. If language ability affects economic outcomes such as lower unemployment and higher earnings, then there are economic incentives for individuals to acquire language skills. As a result language proficiency and earnings might be endogenous.[i]

In the single-equation model there is always a potential bias which could be the result of the endogeneity between earnings levels and English language fluency. Using single equation estimates may produce biased estimates of the true effect of language on earnings. This was first pointed out by McManus *et al*. (1983), but has been the subject of subsequent discussion, notably by Chiswick and Miller (1995), Dustmann and van Soest (1998), and Shields and Price (2001). To investigate this issue two different approaches are adopted. The first involves a simultaneous equation model, whilst the second involves the estimation of a fluency selection model. Such methods have been used successfully for several other countries. Chiswick and Miller (1995) look at male immigrant earnings in the U.S., Canada, Australia and Israel.

The first approach involves estimation of the following simultaneous equation system

$$y_1 = \gamma_1 y_2^* + \alpha_1 X_1 + u_1 \tag{1}$$

$$y_2^* = \gamma_2 y_1 + \alpha_2 X_2 + u_2 \tag{2}$$

and $y_2 = 1$ if $y^*_2 > 0$

$y_2 = 0$ otherwise

where y_1 refers to log of average earnings, y^*_2 refers to English language fluency and X_1 and X_2 contain exogenous variables that influence earnings and fluency. Here only y_1 and y_2 are observed. That is, y^*_2 is only observed as a dichotomous variable y_2. The reduced forms are

$$y_1 = \prod_1 X + v_1 \tag{3}$$

$$y_2^* = \prod_2 X + v_2 \tag{4}$$

where X = [X_1 X_2]. The two structural equations (1) and (2) are estimated in analogous way with two stage least squares. The reduced form earnings equation is estimated by ordinary least squares and the reduced form language equation is estimated by maximum likelihood probit. Following this, the linear predictions from the reduced form equations replace y^*_2 and y_1 on the right hand side of the structural equations (1) and (2) respectively. Since y^*_2 is observed only as a dichotomous variable, the derivation of the covariance matrix follows the method used in Amemiya (1978) for the Nelson-Olsen model.[ii]

For the second approach, a selection model is adopted. Here the earnings equation is estimated separately for fluent and for non-fluent non-whites. This involves estimation of equation (1) whilst dropping y^*_2 from the right hand side and instead selecting on fluency using a selection equation. This is the Heckman (1979) approach selecting on English language fluency.

By running separate equations for each group, the independently developed decomposition method by Blinder (1973) and Oaxaca (1973) can be used.[iii] This decomposes the ethnic wage gap by comparing mean earnings of two groups, namely fluents and non-fluents.[iv] This method leads to two alternative decompositions. The first of which decomposes around fluent average characteristics, whilst the second decomposes around non-fluent average characteristics. The two can give very different results. These are

$$\overline{Y}^F - \overline{Y}^{NF} = \alpha^{NF}(\overline{X}^F - \overline{X}^{NF}) + \overline{X}^F(\alpha^F - \alpha^{NF}) \tag{5}$$

or

$$\overline{Y}^F - \overline{Y}^{NF} = \alpha^F(\overline{X}^F - \overline{X}^{NF}) + \overline{X}^{NF}(\alpha^F - \alpha^{NF}) \quad \textbf{(6)}$$

where *F* refers to fluent and *NF* refers to non-fluent non-whites and therefore the wage offer differential between fluents and non-fluents is $\overline{Y}^F - \overline{Y}^{NF}$ and $\alpha^{NF}(\overline{X}^F - \overline{X}^{NF})$ is the contribution of differences in productivity. $\overline{X}^F(\alpha^F - \alpha^{NF})$ is the contribution of coefficient differences. Productivity differences are those that arise out of differences in characteristics. An example would be differences in levels of qualifications. Since this is a sample of non-whites, split only by language fluency, coefficient differences can be identified as the language effect. In this way it can be seen how non-fluents are additionally penalised in having poorer average characteristics (other than language), which contributes to earnings inequalities.[v]

3. The data.

The *FNSEM* was conducted in 1994 by the Policy Studies Institute.[vi] The survey deliberately oversampled the electoral wards in England and Wales that contain a high percentage of ethnic minorities.[vii] The *FNSEM* contains information on 5196 non-whites and for comparative purposes, 2867 whites aged 16 and over. Six weighting factors were applied to the data to ensure that the survey represented the true populations. These correct for known selection inequalities, observed lack of response rates, to ensure that the sample matched the 1991 Census in terms of age within ethnic group and finally to set the sample size equal to the actual number of interviews.[viii] Non-whites consisted of Black Caribbeans, Indians, Pakistanis, Bangladeshis, African Asians and Chinese.[ix] The definition of ethnicity is based on two questions. The first asks which ethnic group respondents thought they belonged to and the second asks what they considered their family origin to be. The *FNSEM* can therefore distinguish between those who were born abroad, when they arrived and those who were born in Britain.

For our purposes, a key feature of the *FNSEM* is its questions on English language speaking ability. Interviews are conducted in English or in the first language of the respondent. There are three questions concerning the oral and aural English fluency of the respondent. These questions are interviewer assessed, where possible by an interviewer from the same ethnic group. The first question asks the interviewer whether the interview was conducted wholly in English, partially in English or not in English at all. The second question is conditional on the first. If the interview was conducted either partially or wholly in another language, then which other language was used for the interview. The final question asks the interviewer to code the respondent according to their English language ability. The categories of these abilities were fluent, fair, poor or none.

In this study non-fluents consist of those with fair, poor or no language speaking ability. Since the respondent is not asked to provide their own assessment of fluency, there is little chance of self-reported measurement error.[x] According to Sheilds and Price (2001), bias may arise from interviewer error. This would occur should the interviewers systematically over-estimate 'overall' language fluency, on the grounds that they evaluate only English speaking fluency rather than reading and writing skills.

In the *FNSEM* information on annual and weekly average earnings is banded with 16 categories. Since 20 percent of observations having missing earnings information this study uses information on 1324 males and 1219 females only. Furthermore the survey does not contain information on years of schooling or experience. Only information on qualification attainment and age are provided.

[Table 1 here]

Table (1) shows some descriptive statistics concerning the fluency of the sample for whites and non-whites. Males and females are shown separately. As might be expected fluent average weekly earnings are everywhere greater than those for non-fluents and white earnings are everywhere greater than those for non-whites. However the most interesting feature of table (1) concerns females. Unlike males, female average fluent earnings are greater than those for whites. Previous studies have not addressed the issue of language fluency with respect to female earnings. Table (1) suggests that there is no ethnic penalty for females but instead that differences exist between fluents and non-fluents. Table (1) also supports the notion that the gender penalty is much greater than the ethnic penalty for females. This confirms the findings of Leslie *et al* (1998).

4. Earnings Functions for Whites and Non-whites

The analysis begins with estimation of standard earnings functions, correcting for employment selection.[xi] The identification restrictions for the earnings equation are housing tenure, health status, whether has children and whether the individual has a car. Separate functions are estimated for males and females since pooling and including a gender dummy is inadequate if the structural determinants of earnings are gender specific. Weekly earnings are used and these are banded. There are 16 categories and following Chiswick and Miller (1995), the dependent variable is formed from the midpoints of the income intervals and by using a value of 1.5 times the lower threshold.[xii] The specification is typical of this type of work and includes a standard set of independent variables measuring human capital and other socio-economic characteristics. Justification can be found in Leslie (1998). These are age and its square, highest educational qualifications, overseas qualifications, region of residence, marital status, firm size, type of industry and English language non-fluency. Non-fluency consists of those with fair, poor or no language speaking ability. As suggested by Blackaby *et al.* (2001), a measure of the local unemployment rate is also included as well as a distinction between the foreign and native born. Since ethnic minorities are concentrated in local enclaves with high levels of unemployment, the relationship between employment and earnings might be especially significant. Local unemployment rate variables provide the local unemployment rate in the ward that the respondent resides in. Respondents are grouped into one of five local unemployment rate categories; less than 4.99 percent, between 5 and 9.99 percent, between 10 and 14.99 percent, between 15 and 19.99 percent and greater than 20 percent.

Table (2) shows the results for males and females separately. All estimates have been corrected for employment selection. For males, the earnings/age locus displays an inverted U shape and higher educational qualifications imply greater earnings, on average. Relative to those who are married, being single or divorced reduces average earnings of males, although it increases average earnings for females. For males and females, average earnings are significantly higher for those living in London, although they are lower for men living in the midlands, relative to the South of England. Males working in the service industry earn, on average less than those working in the construction industry and working for a firm that employs between 1-49 employees significantly reduces both male and female average earnings.

[Table 2 here]

Table (2) shows that English language fluency does positively influence male earnings. Earnings are 0.111 log points (11.74 percent) less for non-fluent men in the sample. Relative to whites, there is a significant ethnic penalty to foreign born non-whites over and above the language effect of 16.07 percent. This is not so for British born non-white males. The local unemployment rate variables are statistically significant for males. These demonstrate an increasing earnings penalty the greater is the local unemployment rate.

The insignificance of the British born variable supports the idea that the male white/non-white wage gap is closing and that non-white males are assimilating in terms of earnings. Also the significance of the local unemployment rate supports the findings of Blackaby *et al.* (2001). The concentration of ethnic minorities in local enclaves with their high levels of unemployment has a greater influence over the ethnic earnings disadvantage than does English language non-fluency. However the story looks rather different for females.

The second column of table (2) refers to females. One can immediately see that the non-fluency variable is significantly negative. Earnings are 0.283 log points (32.71 percent) less for non-fluent women, relative to whites. This is almost three times more than the fluency penalty for men. The UK born non-white variable is positive and significant and shows that British born non-white women earn approximately 0.173 log points (18.88 percent) more than whites, on average, over and above the language non-fluency penalty. Since all other earnings enhancing characteristics are accounted for, this may well demonstrate positive discrimination for ethnic minority women. After qualifications, it is the language penalty that is stronger for women and it is local unemployment rates that are stronger for males. This result may well reflect differences in the types of employment held by women to those held by men. The supplementary nature of female earnings in terms of contributing to household income may be why local unemployment rates do not impact female weekly earnings.

The final row in table (2) shows the correlation coefficients between the error terms of the employment selection equation and those of the earnings equation These show that the correlation coefficients are significantly negative. So earnings would be higher for those who are unemployed, should they gain employment, relative to those who are already in jobs. Blackaby *et al*. (1999) suggest that this occurs since the unemployed have higher reservation wages than the employed and in turn would require a greater reward to enter into employment.

Table A1, in Appendix A, provides the employment probit estimates. These are consistent with those of previous studies such as Leslie and Lindley (2001), and are therefore only discussed briefly here. As would be expected, age demonstrates an inverted U shape and higher qualifications imply a greater probability of employment. Also being single or divorced implies a lower probability of employment, relative to being married. Living in local or rented accommodation implies a lower propensity for employment than being an owner occupier, whilst living in a household with a car implies higher employment. For males, being in poor health implies a lower probability for employment, as does having children for women. Interestingly, there is a significant employment penalty to non-fluent males, but not to females. Indeed this confirms the results of Leslie and Lindley (2001), where non-fluent women are penalised into economic inactivity rather than unemployment. Finally, there is an employment penalty to British born non-white men and women, over and above lack of language fluency.

5. The Simultaneous Equation Approach

Table (2) made the convenient assumption that language ability is exogenous. Clearly, if language ability affects economic outcomes, then there are economic incentives for individuals to acquire language skills. If the acquisition of language implies higher wages and a higher rate of employment or there is interviewer error, the effect of language on earnings might then be biased.

This simultaneous model begins with the estimation of the reduced form equations (3) and (4). These are provided in Appendix B. The single equation and the simultaneous equation estimates of the structural equations (1) and (2) are presented here. The sample is now restricted to non-whites. Since all whites in the sample are considered to be fluent in English, it is necessary to exclude whites in order to isolate the effects of English language fluency. Again males and females are estimated separately on the basis that the structural determinants of earnings and English language fluency are gender specific.

The specification for the earnings equations is almost identical to that in table (2). Again the dependent variable refers to the log of average earnings taken at the midpoint of each banded category. Also non-whites are disaggregated into their individual ethnic groups, with Black Caribbean being the excluded category rather than whites. Small sample sizes make it necessary to group together Bangladeshi and Pakistani females, females in the construction industry and the production industry, professional and associate professional females, and females with higher and further qualifications. Also ‘immigrant arrival time in Britain’ variables are now included to measure the effects of duration in the host country.[xiii] As a result British born replaces foreign born.

Finally `own ethnic density within ward’ variables are included to measure any neighbourhood effects. Respondents are grouped into one of four categories; 0-5 %, 5-15 %, 15-33 % and above 33%, own ethnic density in the ward of residence. Since local unemployment rates are included as a measure of local economic conditions, the significance of the own ethnic density variables provide evidence of enclave effects on earnings, which are over and above local economic conditions.

This asks whether individuals living in a ward with a high concentration of individuals from their own ethnic group experience significantly different average earnings than those who live in a ward with a low concentration of own ethnic group.

Equation (2) is simply a demand equation for language skills. Its specification is as follows. Age and its square, as well as educational qualifications to try to measure efficiency. To measure exposure, qualifications are split into those attained in Britain and those attained overseas. Also British born, immigrant arrival time, ethnic origin and own ethnic density variables are included. All results are relative to the excluded category of Black Caribbeans (who have a very high proportion of fluents).[xiv]

Also included in equation (2) is a dichotomous variable to measure whether the interview was performed partially in English or alternatively that the interview was conducted wholly in English or wholly in another language.[xv] At the beginning of the interview the respondent is asked whether they would like the interview to be in English or in their own native tongue. If the respondent decides to start the interview in English but then changes to another language, this might be thought to indicate some confidence to converse in English. Since an employer will not have access to the information concerning how respondents behaved in the *FNSEM* but can only observe fluency, this dichotomous variable is used as an exclusion restriction on the earnings equation so that the model can be identified. This follows the work of Shields and Price (2001) who use a similar identifying restriction in their instrumental variable model of occupational success amongst male immigrants. Following Leslie and Lindley (2001), the exclusion restrictions for the language equation are region of residence, marital status and local unemployment rates.[xvi] To facilitate a discussion on the determinants of language fluency, the language probit estimates are also presented.

5.1 Earnings

Table (3) provides the results for the earnings equations for males and females separately. The first column refers to the single equation whilst the second column provides the simultaneous estimates. The earnings/age locus displays an inverted U shape for males. Unlike the single equation estimates, the simultaneous estimates no longer associate higher educational qualifications with greater non-white earnings. Being single significantly reduces the average earnings of males, although marital status has no effect on the average earnings of ethnic minority females. Average earnings are significantly higher for men and women living in the London. For males, 15-33% own ethnic density also implies higher earnings, although they are significantly lower for men working in the service industry. For both men and women working for a small firm significantly reduces average earnings.[xvii] Also high local unemployment rates are significant for males but again for females they are statistically insignificant. The significance of the 15-33 % own ethnic density in ward variable suggests enclave effects upon earnings, which are over and above local unemployment rate effects.

[Table 3 here]

All estimates in table (3) indicate that earnings are significantly lower for non-fluent non-whites. Predicted language non-fluency, derived from the reduced form equation is significant for both males and females. The simultaneous method estimates a non-fluency earnings penalty of 20.0 percentage log points for non-white males and 30.4 percentage log points for non-white females. Hence the single equation estimate underestimates the male simultaneous estimate. This result for males is consistent with the findings of Shields and Price (2001). They estimate an OLS occupational success penalty to non-fluent immigrant males of 10.84 percent, which increases to 18.21 percent when an instrumental variable technique is used. They ascribe this to interview measurement error. For females, the OLS estimate appears to overestimate the female fluency simultaneous estimate. This positive endogenous bias is intuitive if there are rewards associated with the acquisition of language fluency.

Relative to Black Caribbeans, the simultaneous estimates show no significant ethnic penalty over and above the language penalty to the other ethnic groups. This suggests no variations between ethnic groups, once language fluency has been taken into consideration. The single equation estimates in table (3) indicate a significant earnings penalty to immigrant non-white males relative to native born non-white males. However the simultaneous estimates show no significant difference between the British born and those born overseas. Hence nothing here indicates that British born non-whites are doing better in terms of earnings than immigrants. What are more important for males are small firm size, type of industry, whether the individual resides in London and high local unemployment rates.

5.2 Language Determinants

Table (4) provides estimates of the structural language probits associated with table (3). The first row shows that predicted average earnings derived from the reduced form equations are insignificant for both males and females. Higher British qualifications significantly reduce the propensity to be non-fluent for both males and females. Unlike males, females working in the service industry are more likely to be fluent, relative to those who are employed in the production or construction industry. The 15 to 33 percent own ethnic density variable does significantly negatively impact the fluency of men, relative to those in low own ethnic density areas. This supports the idea that minorities in areas of high ethnic concentration should be less proficient in the host language. According to Chiswick and Miller (1995), immigrants in areas of high ethnic concentration should be less proficient in the host language simply because there is less necessity to acquire language skills.

[Table 4 here]

Not surprisingly the ‘interview partially in English’ variable has a positive relationship with non-fluency and British born non-whites have a lower propensity to be non-fluent. However the immigrant arrival time variables seem to explain very little in terms of non-fluency. What seems more important is ethnicity. All males have a higher propensity to be non-fluent relative to Black Caribbean males, with Pakistanis and Indians displaying the highest propensity for non-fluency. All females have a higher propensity to be non-fluent relative to Black Caribbeans. Indians and Chinese display the highest non-fluency, with African Asian women displaying the lowest. Compared with males, the ethnic penalties for female non-fluency appear to be generally bigger. As with males, the British born have a significantly lower propensity to be non-fluent relative to immigrants. Unlike males, immigrant females from the 1960’s cohort are also less likely to be fluent. This may well reflect increases in South Asian immigrants typical of that time. This could be associated with cultural differences between cohorts, since ethnicity is held constant.[xviii]

Overall, the structural approach suggests a bigger role for fluency impacting on earnings. However, it may be that the results here are sensitive to the choice of identification instruments. According to Bound *et al*. (1995) there is a possibility of large biases when the identification instruments have only a weak influence on the endogenous explanatory variable. Appendix B contains F-tests on the reduced form earnings equations with and without the excluded language instrument. These show no evidence of under-identification. For the language equations, Appendix B performs Likelihood Ratio tests on the reduced form probits. Again these show no evidence of under-identification.

6. The Selection Equation Approach

The selection equation approach involves the estimation of equation (1) whilst dropping y^{*}_{2} from the right hand side and instead selecting on fluency using a selection equation. The specification for this selection equation is identical to the language probit in the previous section although average earnings are now dropped from the right hand side. The sample again consists of non-whites only. Again, `whether the interview was conducted partially in English' is used to identify the earnings equation. Finally, males and females are pooled here as a result of the small sample numbers for non-fluents. As a result the earnings equation now includes a dichotomous variable intended to capture gender differences.

Table (5) provides the results for three equations. The first column refers to the single equation OLS estimate for the pooled sample of fluents and non-fluents. The second column refers to the selection equation, selecting on fluents. The third column refers to a selection equation selecting on non-fluents. The final row in table (5) shows the correlation coefficients between the error terms of the selection equation and those of the earnings equation.

[Table 5 here]

Comparing the first column with the second shows the differences between the pooled sample and fluents. Amongst those fluent in English, the partial effect of UK education on earnings is lower than those reported in the pooled analysis (for example higher qualification this is 39.1 percentage log points compared with 94.2). Also overseas qualifications imply higher earnings for fluents. Comparing the first and the third columns show the differences between the pooled sample and non-fluents. The partial effect of being female compared to being male is much more negative for non-fluents than for the aggregate-level result (-50.6 percentage log points compared with -35.8). Hence non-fluent females appear to suffer disproportionately lower earnings than fluent females. Also qualifications have no significant effect on the earnings of non-fluents, even though the effect of qualifications on fluent earnings are positive and significant. This is consistent with Chiswick and Miller (1995) and suggests some complementarity between the skills represented by formal education and language.[xix]

Interestingly, ethnicity and immigrant status play no part in determining non-fluent earnings. However, for fluents there is an earnings penalty only to foreign born Bangladeshis. This further suggests no earnings penalty to the British born once fluency is controlled for. Most of the earnings differences between fluents and non-fluents are explained by differences in qualifications. The insignificance of the correlation coefficients demonstrates that the fluency selection equations are insignificant. Hence there is no evidence here that language fluency is endogenous to the process of earnings determination amongst Britain's ethnic minorities and demonstrates the problems associated with modelling such endogenous relationships.

[Table 6 here]

Table (6) provides the Blinder/Oaxaca decompositions between fluents and non-fluents. These are a decomposition of the total differences in earnings between fluents and non-fluents. Here two single equation earnings functions are estimated, one for fluents and another for non-fluents. These are then decomposed. On the basis of the results in table (5), no fluency selection is assumed.

In table (6) the first row provides the total average pay differential, whilst the second and third decompose this into coefficient effects and the characteristic effects respectively. The first row of table (6) confirms that non-fluent non-whites are disadvantaged relative to fluent non-whites in terms of average earnings. Focussing on the actual decompositions, differences in fluent/non-fluent average earnings can be mainly ascribed to differences in other (non-language) characteristics (around 58 percent of the total differential). These include education, ethnicity and whether or the individual was born in the UK. However, there is a strong language effect here of around 40 percent of the total difference between fluent and non-fluent earnings.

7. Concluding comments.

This study provides an analysis of the earnings gap between fluent and non-fluent ethnic minorities as well as the determinants of English language fluency. Lack of fluency in English has a significant impact on the average earnings of both ethnic minority men and women in Britain. Moreover the language penalty for women is almost three times larger than it is for males. There is no earnings penalty to British non-white males above language fluency. This suggests that ethnic minority earnings are assimilating towards those of whites. In addition, British born non-white women earn more than white women, on average. Hence it is poor language fluency that is the issue for non-white women, rather than ethnicity.

In comparison to the simultaneous approach, the single equation estimates appear to underestimate the true language penalty for males and overestimate the language penalty for females. Indeed the simultaneous estimates indicate no average wage differential between native born and immigrant ethnic minorities and that no variations exist between ethnic groups. This further confirms assimilation. For men, type of industry, employment with a small firm, high local unemployment rates and non-fluency in English are the main determinants of non-white earnings. For women it is employment within a small firm and English language non-fluency.

The main determinants of English language fluency are British and overseas qualifications, immigrant status and ethnic origin. Results imply higher fluency for increased efficiency and greater exposure. British born ethnic minorities have a lower propensity to be non-fluent. However it is evident that there are significant variations across Britain's ethnic groups. All employed ethnic minorities have a higher propensity to be non-fluent relative to Black Caribbeans, with Pakistani males, and Indian and Chinese females exhibiting the highest probabilities of non-fluency.

Comparisons of the coefficients on the qualifications variable in samples of fluents and non-fluents indicate complementarity among skills. Therefore a policy that aimed to get non-whites into further education would both improve their fluency and qualifications, and have a dual impact on ethnic minority earnings. However, the taxation costs associated with lower fluency are small. Predicted average gross weekly earnings in the sample were £259.02 (£189.29) for men (women). If everyone in the sample became fluent then predicted average earnings would have increased to £266.62 (£204.64) for men and women. Assuming an approximate income tax rate of 20 percent, then the average non-fluent man (woman) contributed £79.04 (£159.64) less income tax in 1994. According to the *LFS,* ethnic minorities constituted only 5.9 percent of the British working population in 1994. So the numbers of non-fluents would have been relatively small. Hence the costs of fluency in terms of tax revenue would have been negligible.

If higher earnings can be associated with improved fluency, this provides some incentive for the disadvantaged to bear the cost of language acquisition. Since lack of fluency can be associated with the foreign born, and a more education implies greater fluency, then one might speculate that fluency is assimilating with each passing generation. Making language proficiency a pre-requisite for entry should speed up any future assimilation process.

References.

Amemiya, T. (1978), "The Estimation of a Simultaneous Equation Generalised Probit Model", *Econometrica*, Vol. 46, No. 5, pp. 1193-1205.

Bell, B.D. (1997), "The Performance of Immigrants in the United Kingdom: Evidence from the GHS," *The Economic Journal*, Vol. 107, pp. 333-44.

Blackaby, D.H., Clark, K., Leslie D.G. and Murphy, P.D. (1998), "The Ethnic Wage Gap and Employment Prospects in the 1990s: Evidence for Britain," *Economics Letters*, Vol. (58), pp. 97-103.

Blackaby, D.H., Clark, K., Leslie D.G. and Murphy, P.D. (1994), "Black-White Earnings and Employment Prospects in the 1970s and 1980s: Evidence for Britain," *Economics Letters*, Vol. (46), pp. 273-79.

Blackaby, D., Drinkwater, S., Murphy, P. and O'Leary, N. (2001), "English Language Fluency and the Ethnic Wage Gap in Britain", forthcoming*, Economic Issues.*

Blackaby, D., Leslie, D., Murphy, P. and O'Leary, N. (1999) "White/Ethnic Minority Earnings and Employment Differentials in Britain: Evidence from the *LFS*," *MIMEO,* University of Wales, Swansea.

Blinder, A.S. (1973), "Wage Discrimination: Reduced Form and Structural Variables", *Journal of Human Resources,* Vol 8, pp. 436-55.

Borjas, G.J. (1994), "The Economics of Immigration", *Journal of Economic Literature,* Vol. 37, pp. 1667-1717.

Chiswick, R.B. (1978), "The Effect of Americanisation on the Earnings of Foreign Born Men", *Journal of Political Economy,* Vol. 86, pp. 897-921.

Chiswick, B.R. and Miller, P.W. (1995), "The Endogeneity between Language and Earnings: International Analyses", *Journal of Labor Economics,* 13(2), pp. 246-288.

Dustmann, C. and van Soest, A. (1998), "Language and the Earnings of Immigrants". *Mimeo,* University College London: September.

Heckman, J. (1979), "Sample Selection Bias as a Specification Error", *Econometrica*, Vol. 47, pp. 153-61.

Leslie, D. (ed.) (1998), *An Investigation of Racial Disadvantage,* Manchester University Press, Manchester.

Leslie, D., Drinkwater, S., and O'Leary N. (1998), "Unemployment and Earnings among Britain's Ethnic Minorities: Some Signs for Optimism" , *Journal of Ethnic and Migration Studies*, Vol. 24, pp. 489-506

Leslie, D. and Lindley, J.K., (2001), "The Impact of Language Ability on Employment and Earnings of Britain's Ethnic Communities, Forthcoming, *Economica*.

McManus, W., Gould W. and Welch (1983), "Earnings of Hispanic men: the role of English language proficiency", Journal of Labor Economics, 1, pp. 101-130.

Miller, P.W., (1989), "An Economic Analysis of Citizenship in Australia", *Nedlands*: University of Western Australia, Department of Economics, 1989.

Modood, T., Berthoud, R. and others (1997) *Ethnic Minorities in Britain: Diversity and Disadvantage*, Policy Institute, London.

Neumark, D. (1988), “Employers Discriminatory Behaviour and the Estimation of Wage Discrimination”, *Journal of Human Resources,* Vol 23, pp. 279-95.

Oaxaca, R.L., (1973), “Male-Female Wage Differentials in Urban Labour Markets”, *International Economic Review,* Vol 14, pp.529-36.

Oaxaca, R.L. and Ransom, M.R. (1994), “On Discrimination and the Decomposition of Wage Differentials”, *Journal of Econometrics,* Vol 61, pp.5-21.

Shields, M.A. and Wheatley Price, S. (2001), “The English language fluency and occupational success of ethnic minority immigrant men living in English metropolitan areas”, Forthcoming, *Journal of Population Economics*.

Smith, P. and Prior, G. (1996), *The Fourth National Survey of Ethnic Minorities: Technical Report,* London: Social and Community Planning Research.

Stromback, T., (1986), “The Earnings of Migrants in Australia”, Conference Paper No 46, Canberra: *Bureau of Labour Market Research.*

Table (1) Average weekly earnings in pounds sterling, (1994 FNSEM).

	Males	Females
Non-white fluents.	302.11 (559)	231.36 (505)
Non-white non-fluents.	178.03 (235)	140.15 (124)
Whites.	345.90 (530)	196.91 (590)

Notes: Figures in parentheses are sample sizes.

Table (2) Earnings functions for whites and non-whites corrected for employment selectivity, (1994 FNSEM). (Dependent Variable log of average Earnings).

	Males	Females
Age	0.059 (3.76)*	0.021 (1.68)
Age squared	-0.001 (3.15)*	-0.001 (1.58)
Highest qual: higher	0.593 (10.54)*	0.702 (8.18)*
Highest qual: further	0.285 (4.69)*	0.445 (5.42)*
Highest qual: olevh	0.231 (4.32)*	0.254 (3.53)*
Highest qual: othqual	0.101 (1.94)*	0.177 (2.35)*
Quals attained abroad.	0.127 (2.38)*	0.327 (3.09)*
Single	-0.114 (2.59)*	0.217 (4.15)*
Divorced.	-0.102 (1.37)	0.186 (2.79)*
Lives in the North of England.	-0.076 (1.76)	-0.061 (0.93)
Lives in the Midlands.	-0.097 (2.39)*	-0.038 (0.58)
Lives in Wales.	-0.052 (0.44)	-0.126 (0.65)
Lives in London.	0.117 (2.61)*	0.293 (4.65)*
Production Industry	-0.093 (1.20)	-0.297 (1.06)
Service Industry	-0.162 (2.09)*	-0.341 (1.24)
Firm size: 1-49 emps	-0.261 (7.51)*	-0.257 (5.08)*
Firm size: 50-99 emps	-0.089 (1.84)	-0.046 (0.66)
Firm size: 100-249 emps	-0.029 (0.73)	-0.001 (0.02)
Firm size: 250-499 emps	-0.013 (0.27)	-0.082 (1.01)
Non-white (UK born)	-0.064 (1.29)	0.173 (2.44)*
Non-white (Foreign born)	-0.149 (2.97)*	-0.011 (0.15)
Non-Fluency	-0.111 (2.07)*	-0.283 (3.09)*
Unemployment rate in local ward.		
5-9.99 %	-0.116 (2.41)*	-0.059 (0.77)
10-14.99 %	-0.193 (3.67)*	-0.036 (0.43)
15-19.99 %	-0.296 (4.74)*	0.027 (0.29)
>20 %	-0.289 (4.78)*	0.089 (0.95)
Constant	4.694 (14.75)*	4.821 (12.31)*
Correlation coefficient, ρ	-0.669 (4.47)*	-0.758 (7.49)*
Sample size	1324	1219

*Notes: t statistics in parentheses * denotes statistical significance at the 5% level.*
The excluded category are married whites, with no qualifications, in good health, working for a firm with more than 500 employees, who works in the construction industry and lives in a low unemployment area in the South of England(excluding London).

Table (3), Earnings functions for non-whites only (1994 FNSEM). (Dependent Variable log of average Earnings).

	Males		Females	
	Single Equation	Simultaneous Equation	Single Equation	Simultaneous Equation
Non Fluent [a]	-0.176 (3.47)*	-0.200 (3.08)*	-0.334 (3.55)*	-0.304 (2.39)*
Age	0.048 (4.52)*	0.050 (3.46)*	0.043 (2.16)*	0.021 (0.71)
Age squared	-0.001 (4.15)*	-0.001 (3.13)*	-0.001 (1.82)	-0.001 (0.39)
Highest qual: higher	0.706 (9.76)*	0.221 (0.99)	-	-
Highest qual: further	0.180 (2.52)*	-0.170 (1.01)	-	-
Highest qual: higher or further	-	-	0.789 (7.73)*	0.258 (0.88)
Highest qual: olevh	0.194 (2.85)*	-0.016 (0.13)	0.404 (3.87)*	-0.046 (0.18)
Highest qual: othqual	0.130 (1.98)	-0.083 (0.69)	0.428 (4.23)*	-0029 (0.11)
Qualifications attained abroad	0.169 (2.87)*	0.113 (1.58)	0.351 (3.31)*	0.238 (1.71)
Single	-0.159 (2.52)*	-0.233 (2.63)*	-0.055 (0.74)	-0.069 (0.59)
Divorced	-0.185 (1.41)	-0.176 (1.11)	0.004 (0.05)	0.045 (0.34)
Lives in the North of England.	-0.048 (0.68)	0.094 (0.89)	-0.130 (1.15)	0.003 (0.02)
Lives in the Midlands.	-0.107 (1.71)	0.043 (0.43)	0.070 (0.75)	0.390 (2.00)*
Lives in Wales.	0.166 (1.08)	0.331 (1.65)	-0.166 (0.56)	0.836 (0.18)
Lives in London.	0.092 (1.60)	0.205 (2.39)*	0.288 (3.52)*	0.307 (2.67)*
Production Industry	-0.248 (2.12)*	-0.261 (1.75)	-	-
Service Industry	-0.303 (2.77)*	-0.362 (2.47)*	-0.161 (0.45)	-0.205 (1.72)
Firm size: 1-49 employees	-0.301 (6.06)*	-0.351 (5.49)*	-0.145 (1.99)*	-0.214 (2.07)
Firm size: 50-99 employees	-0.218 (3.13)*	-0.371 (3.53)*	-0.033 (0.33)	-0.024 (0.16)
Firm size: 100-249 employees	-0.074 (1.20)	-0.147 (1.82)	0.153 (1.47)	-0.012 (0.07)
Firm size: 250-499 employees	-0.097 (1.36)	-0.167 (1.78)	0.007 (0.06)	-0.102 (0.61)
British born	0.203 (2.21)*	0.003 (0.03)	-0.123 (0.71)	-0.509 (1.76)
Arrived pre 1959	-0.067 (0.49)	0.140 (0.80)	-0.165 (0.65)	-0.293 (0.83)
Arrived 1960-1969	0.191 (2.19)*	0.085 (0.75)	-0.164 (0.95)	-0.423 (1.71)
Arrived 1970-1979	0.182 (2.25)*	0.095 (0.93)	-0.146 (0.88)	-0.191 (0.96)
Arrived 1980-1989	0.043 (0.49)	-0.121 (0.11)	-0.183 (1.02)	-0.290 (1.34)
African Asian	-0.112 (1.68)	0.035 (0.33)	-0.001 (0.01)	0.349 (1.58)
Indian	-0.085 (1.42)	0.152 (1.27)	-0.047 (0.56)	0.471 (1.64)
Pakistani	-0.111 (1.54)	0.125 (0.96)	-	-
Bangladeshi	-0.306 (3.72)*	-0.059 (0.42)	-	-
Pakistani & Bangladeshi	-	-	-0.153 (1.27)	0.265 (0.99)
Chinese	-0.140 (1.56)	0.091 (0.61)	0.036 (0.29)	0.471(1.61)
15-33 % own ethnic density	0.052 (1.04)	0.149 (2.19)*	-0.071 (0.92)	0.053 (0.13)
> 33 % own ethnic density	0.006 (0.09)	-0.015 (0.17)	-0.155 (1.04)	0.108 (0.49)
Unemployment rate in local ward:				
5-9.99 %	-0.101 (1.09)	-0.104 (0.75)	0.076 (0.48)	0.088 (0.40)
10-14.99 %	-0.205 (2.23)*	-0.197 (1.44)	0.143 (0.89)	0.203 (0.90)
15-19.99 %	-0.364 (3.56)*	-0.295 (1.98)*	0.080 (0.46)	0.100 (0.41)
>20 %	-0.349 (3.49)*	-0.363 (2.52)*	0.079 (0.45)	0.036 (0.15)
Constant	4.788 (17.55)*	4.618 (11.81)*	4.00 (6.80)*	3.984(5.98)*
R Squared	0.46	0.47	0.27	0.27
Sample size	794	794	629	629

Notes: a This is predicted non-fluent in the simultaneous equation case. t statistics in parentheses.
** denotes statistical significance at the 5% level..*
The excluded category are married Black Caribbeans, with no qualifications, arrived in Britain between 1990 and 1994, who works for a firm with more than 500 employees and works in the construction industry, living in a low unemployment and low ethnic density area outside the South of England (excluding London).

Table (4). Structural language probits for non-whites only (1994 FNSEM).
(Dependent Variable=1 if non-fluent, 0 if fluent).

	Males	Females
Log of average earnings. [a]	-0.602 (1.27)	-0.806 (1.27)
Age.	0.088 (1.63)	0.011 (0.16)
Age Square.	-0.001 (1.60)	-0.001 (0.03)
Highest qual: higher	-2.097 (3.15)*	-
Highest qual: further	-1.776 (4.76)*	-
Highest qual: higher or further	-	-1.156 (1.78)*
Highest qual: olevh	-1.207 (4.24)*	-1.203 (2.68)*
Highest qual: othqual	-1.230 (4.58)*	-1.256 (2.81)*
Qualifications attained abroad.	0.283 (1.37)	-0.268 (0.75)
Married to a UK born spouse	0.432 (1.47)	-0.657 (1.17)
Production Industry	0.056 (0.11)	-
Service Industry	-0.210 (0.42)	-0.612 (2.91)
Lives in LA accommodation	-0.243 (0.97)	-0.380 (0.25)
Lives in rented accommodation	-0.228 (0.84)	-0.306 (0.77)
Lives in other accommodation	0.156 (0.28)	-0.275 (0.27)
Firm size: 1-49 employees	-0.338 (1.40)	-0.423 (1.75)
Firm size: 50-99 employees	-0.845 (2.75)*	-0.316 (0.88)
Firm size: 100-249 employees	-0.219 (0.87)	-0.430 (1.13)
Firm size: 250-499 employees	-0.452 (1.59)	-0.564 (1.47)
Interview undertaken partially in English.	0.794 (4.21)*	0.695 (2.36)*
British born.	-1.069 (2.95)*	-1.245 (2.66)*
Arrived pre 1959	-0.606 (1.19)	-0.803 (0.99)
Arrived 1960-1969	-0.518 (1.75)	-1.071 (2.39)*
Arrived 1970-1979	-0.495 (1.79)	-0.497 (1.24)
Arrived 1980-1989	-0.324 (1.16)	-0.602 (1.40)
African Asian	0.649 (2.33)*	1.075 (2.96)*
Indian	1.147 (4.35)*	1.646 (5.03)*
Pakistani	1.277 (4.52)*	-
Bangladeshi	0.920 (2.73)*	-
Pakistani & Bangladeshi	-	1.213 (2.81)*
Chinese	0.100 (2.47)*	1.411 (3.02)*
15-33 % own ethnic density	0.493 (3.10)*	0.296 (1.22)
> 33 % own ethnic density	-0.163 (0.72)	0.577 (1.61)
Constant	0.741 (0.31)	3.636 (1.37)
Pseudo R^2	0.47	0.52
Sample size	794	629

Notes: *a This is predicted log of average earnings taken from the earnings equation.*
** denotes statistical significance at the 5% level.*
The excluded category are Black Caribbeans, with no qualifications, who arrived in Britain between 1990 and 1994, living in a low ethnic density area in the South of England (excluding London).
t statistics in parentheses

Table (5) Results for earnings functions selecting on fluents and non-fluents (1994 FNSEM). (Dependent Variable log of average Earnings).

	OLS	Selecting on Fluents	Selecting on Non-Fluents
Female	-0.358 (10.00)*	-0.294 (7.58)*	-0.506 (6.03)*
Age	0.043 (3.38)*	0.052 (3.44)*	0.031 (1.16)
Age squared	-0.001 (2.91)*	-0.001 (2.88)*	-0.001 (1.09)
Highest qual: higher	0.942 (17.07)*	0.391 (4.31)*	0.403 (1.75)
Highest qual: further	0.461 (7.41)*	0.300 (3.58)*	0.154 (0.81)
Highest qual: olevh	0.344 (6.02)*	0.255 (3.09)*	-0.094 (0.56)
Highest qual: othqual	0.324 (5.49)*	0.218 (2.49)*	-0.053 (0.23)
Overseas Qualification	0.257 (4.62)*	0.351 (4.70)*	0.136 (1.63)
Single	-0.098 (2.11)*	-0.103 (2.01)*	-0.066 (0.46)
Divorced.	-0.062 (0.76)	-0.092 (1.08)	-0.059 (0.32)
Lives in the North of England.	-0.075 (1.06)	-0.063 (0.79)	-0.098 (0.66)
Lives in the Midlands.	-0.052 (0.91)	-0.074 (1.18)	0.118 (0.89)
Lives in Wales.	0.067 (0.52)	-0.085 (0.53)	0.196 (0.99)
Lives in London.	0.197 (3.83)*	0.212 (3.87)*	0.099 (0.79)
Production Industry	-0.272 (3.04)*	-0.205 (2.09)*	-0.628 (3.28)*
Construction Industry	-0.313 (3.61)*	-0.339 (3.61)*	-0.656 (3.34)*
Firm size: 1-49 emps	-0.231 (5.46)*	-0.289 (6.40)*	-0.142 (1.71)
Firm size: 50-99 emps	-0.104 (1.81)	-0.217 (3.51)*	0.182 (1.18)
Firm size: 100-249 emps	0.005 (0.11)	-0.104 (1.82)	0.209 (1.83)
Firm size: 250-499 emps	-0.024 (0.42)	-0.116 (1.82)	0.118 (1.00)
British born	0.178 (2.06)*	0.102 (0.72)	0.119 (0.73)
Arrived pre 1959	0.044 (0.35)	-0.049 (0.29)	-0.189 (0.64)
Arrived 1960-1969	0.163 (1.89)	0.073 (0.51)	-0.090 (0.76)
Arrived 1970-1979	0.125 (1.52)	0.086 (0.63)	0.113 (1.02)
Arrived 1980-1989	-0.008 (0.09)	0.012 (0.08)	-0.020 (0.17)
African Asian	-0.094 (1.74)	-0.046 (0.77)	0.062 (0.27)
Indian	-0.151 (2.91)*	-0.004 (0.63)	-0.052 (0.23)
Pakistani	-0.147 (2.23)*	-0.042 (0.47)	0.029 (0.13)
Bangladeshi	-0.336 (4.17)*	-0.323 (2.90)*	0.071 (0.29)
Chinese	-0.147 (1.93)	-0.103 (1.21)	0.103 (0.39)
15-33 % own ethnic density	-0.023 (0.55)	0.059 (1.17)	-0.092 (1.13)
> 33 % own ethnic density	-0.108 (1.57)	-0.159 (1.73)	0.088 (0.80)
5-9.99 % local unemployment	-0.071 (0.76)	-0.084 (0.90)	0.323 (1.06)
10-14.99 % local unemployment	-0.091 (0.98)	-0.104 (1.15)	0.330 (1.11)
15-19.99 % local unemployment	-0.192 (1.91)	-0.195 (1.88)	0.246 (0.83)
>20 % local unemployment	-0.189 (1.90)	-0.184 (1.83)	0.152 (0.51)
Constant	4.66 (15.87)*	4.61 (13.72)*	4.61 (7.07)*
Correlation coefficient, ρ	-	-0.243 (1.24)	0.318 (1.58)
Sample size	1423	1423	1423
Censored obs	-	359	1064
Uncensored obs	-	1064	359

Notes: *The excluded category are married Black Caribbean males, with no qualifications, in good health, arrived in Britain between 1990 and 1994, who works for a firm with more than 500 employees, who works in the construction industry, living in a low unemployment and low ethnic density area outside the South of England (excluding London).*
*t statistics are in parentheses. * denotes statistical significance at the 5% level.*

<u>Table (6) Average earnings Decompositions for language fluency for non-whites (1994 FNSEM).</u>

	Males and Females
Differences in Means	
$\overline{Y}^{F} - \overline{Y}^{NF}$	0.485
Differences in Coefficients	
$\overline{X}^{F}(a^{F} - a^{NF})$	0.210
$\overline{X}^{NF}(a^{F} - a^{NF})$	0.204
Differences in Characteristics	
$a^{NF}(\overline{X}^{F} - \overline{X}^{NF})$	0.275
$a^{F}(\overline{X}^{F} - \overline{X}^{NF})$	0.281

Appendix A: Sample selection employment probits.

Table (A.1) Employment probits for whites and non-whites, (1994 FNSEM). (Dependent Variable=1 if employed and 0 if unemployed).

	Males	Females
Age	0.043 (1.96)*	0.153 (6.45)*
Age squared	-0.001 (1.97)*	-0.001 (5.55)*
Highest qual: higher	0.543 (3.63)*	0.725 (3.74)*
Highest qual: further	0.312 (2.50)*	0.500 (3.13)*
Highest qual: olevh	0.132 (1.18)	0.413 (3.19)*
Highest qual: othqual	0.120 (1.09)	0.381 (2.96)*
Quals from abroad.	-0.067 (0.62)	0.175 (0.99)
Single	-0.269 (2.39)*	-0.516 (5.04)*
Divorced.	-0.333 (2.15)*	-0.304 (2.39)*
Lives in the North	-0.276 (1.78)	0.066 (0.49)
Lives in the Midlands.	-0.073 (0.76)	-0.059 (0.45)
Lives in Wales.	-0.121 (0.50)	-0.259 (0.96)
Lives in London.	-0.028 (0.28)	-0.243 (1.99)*
Lives in LA accom.	-0.683 (7.88)*	-0.235 (2.32)*
Lives in rented accom.	-0.452 (4.81)*	-0.293 (2.89)*
Lives in other accom.	-0.473 (3.12)*	-0.402 (1.66)
Non-white (UK born)	-0.495 (4.51)*	-0.279 (2.38)*
Non-white (Foreign born)	0.145 (1.32)	-0.070 (0.60)
Non-Fluency	-0.425 (4.49)*	-0.056 (0.40)
Has children	-0.015 (0.16)	-0.379 (3.82)*
In poor health	-0.512 (4.60)*	0.199 (1.35)
Lives in household with car.	0.556 (6.98)*	0.464 (5.73)*
Constant	0.061 (0.14)	-2.028 (4.45)*
Sample size	1929	1531

*Notes: t statistics in parentheses. * denotes statistical significance at the 5% level.*
The excluded category are married whites, with no qualifications, in good health, with no children, is an owner occupier, without access to a car, working for a firm with more than 500 employees, who works in the construction industry and lives in a low unemployment area in the South of England(excluding London).

Appendix B: Testing the validity of the instruments in the simultaneous equation estimates.

Table (B.1). Reduced form earnings functions for non-whites only (1994 FNSEM).
(Dependent Variable log of average Earnings).

	Males	Females
Age	0.048 (4.55)*	0.041 (2.05)*
Age squared	-0.001 (4.12)*	-0.001 (1.75)
Highest qual: higher	0.756 (10.98)*	-
Highest qual: further	0.226 (3.31)*	-
Highest qual: higher or further	-	0.884 (9.08)*
Highest qual: olevh	0.248 (3.77)*	0.492 (4.87)*
Highest qual: othqual	0.175 (2.76)*	0.505 (5.19)*
Qualifications attained abroad	0.181 (3.08)*	0.378 (3.57)*
Single	-0.138 (2.18)*	-0.059 (0.81)
Divorced	-0.208 (1.58)	0.009 (0.09)
Lives in the North of England.	-0.056 (0.97)	-0.136 (1.20)
Lives in the Midlands.	-0.117 (1.87)	0.053 (0.57)
Lives in Wales.	0.127 (0.83)	-0.183 (0.62)
Lives in London.	0.086 (1.52)	0.298 (3.63)*
Production Industry	-0.205 (2.23)*	-
Service Industry	-0.302 (2.76)*	-0.050 (0.65)
Firm size: 1-49 employees	-0.322 (6.36)*	-0.122 (1.68)
Firm size: 50-99 employees	-0.224 (3.18)*	-0.051 (0.52)
Firm size: 100-249 employees	-0.112 (1.79)	0.156 (1.49)
Firm size: 250-499 employees	-0.106 (1.47)	0.275 (0.24)
British born	0.227 (2.49)*	-0.064 (0.37)
Arrived pre 1959	-0.028 (0.20)	-0.095 (0.37)
Arrived 1960-1969	0.213 (2.46)*	-0.079 (0.46)
Arrived 1970-1979	0.215 (2.66)*	-0.091 (0.55)
Arrived 1980-1989	0.075 (0.85)	-0.148 (0.83)
African Asian	-0.104 (1.55)*	-0.006 (0.06)
Indian	-0.081 (1.36)	-0.091 (1.09)
Pakistani	-0.134 (1.89)	-
Bangladeshi	-0.292 (3.52)*	-
Pakistani & Bangladeshi	-	-0.163 (1.35)
Chinese	-0.150 (1.67)	-0.047 (0.39)
15-33 % own ethnic density	0.047 (0.96)	-0.065 (0.84)
> 33 % own ethnic density	-0.004 (0.05)	-0.168 (1.15)
Unemployment rate in local ward:		
5-9.99 %	-0.094 (1.02)	0.067 (0.43)
10-14.99 %	-0.196 (2.12)*	0.119 (0.74)
15-19.99 %	-0.357 (3.48)*	0.052 (0.30)
>20 %	-0.332 (3.32)*	0.068 (0.39)
Interview undertaken partially in English.	-0.182 (3.63)*	-0.275 (2.91)*
Constant	4.781 (16.59)*	3.812 (8.03)*
R Squared	0.46	0.28
F-statistic [a]	13.24 [3.84]	8.45 [3.84]
Sample size	794	629

Notes: *t statistics in parentheses.* ** denotes statistical significance at the 5% level.*
a F test for the significance of the identification restriction. F-values at the 5% level are in square brackets.
The excluded category are married Black Caribbeans, with no qualifications, in good health, arrived in Britain between 1990 and 1994, who works for a firm with more than 500 employees and is an owner occupier, living in a low unemployment and low ethnic density area outside the South of England (excluding London).

Table (B.2). Reduced form language probits for non-whites only (1994 FNSEM). (Dependent Variable=1 if non-fluent, 0 if fluent).

	Males	Females
Age.	0.009 (0.19)	-0.065 (0.94)
Age Square.	-0.001 (0.24)	0.001 (1.13)
Highest qual: higher	-2.673 (4.84)*	-
Highest qual: further	-1.984 (5.52*	-
Highest qual: higher or further	-	-2.056 (5.94)*
Highest qual: olevh	-1.316 (5.17)*	-1.769 (5.21)*
Highest qual: othqual	-1.291 (5.15)*	-1.758 (5.09)*
Qualifications attained abroad.	-0.333 (1.88)*	-0.461 (1.72)
Production Industry	-0.053 (0.11)	-
Service Industry	-0.303 (0.64)	-0.508 (2.41)*
Firm size: 1-49 employees	-0.141 (0.76)	-0.301 (1.30)
Firm size: 50-99 employees	-0.733 (2.52)	-0.246 (0.69)
Firm size: 100-249 employees	-0.177 (0.72)	-0.553 (1.51)
Firm size: 250-499 employees	-0.307 (1.10)	-0.425 (1.10)
Interview undertaken partially in English.	0.912 (5.81)*	0.904 (3.72)*
British born.	-1.117 (3.19)*	-1.465 (3.10)*
Arrived pre 1959	-0.559 (1.09)	-0.652 (0.79)
Arrived 1960-1969	-0.637 (2.29)*	-1.133 (2.57)*
Arrived 1970-1979	-0.592 (2.33)*	-0.328 (0.83)
Arrived 1980-1989	-0.433 (1.58)	-0.467 (1.12)
African Asian	0.695 (2.44)*	1.129 (2.95)*
Indian	1.172 (4.40)*	1.848 (5.38)*
Pakistani	1.297 (4.57)*	-
Bangladeshi	1.162 (3.68)*	-
Pakistani & Bangladeshi	-	1.407 (3.21)*
Chinese	1.205 (3.02)*	1.707 (3.82)*
15-33 % own ethnic density	0.510 (2.97)*	0.264 (1.03)
> 33 % own ethnic density	-0.096 (0.40)	0.909 (2.33)*
Lives in the North of England.	0.755 (2.62)*	0.457 (1.21)
Lives in the Midlands.	0.799 (3.00)*	1.107 (3.39)*
Lives in Wales.	1.018 (1.78)*	0.327 (0.27)
Lives in London.	0.594(2.37)*	0.029 (0.11)
Single	-0.478 (1.67)	-0.031 (0.10)
Divorced	0.159 (0.36)	0.116 (0.38)
Unemployment rate in local ward:		
5-9.99 %	-0.048 (0.09)	0.067 (0.13)
10-14.99 %	-0.005 (0.01)	0.275 (0.51)
15-19.99 %	0.310 (0.59)	0.155 (0.26)
>20 %	-0.153 (0.29)	-0.105 (0.18)
Constant	-0.842 (0.66)	0.565 (0.35)
Pseudo R^2	0.52	0.58
Likelihood Ratio Test	20.67 [18.31]	22.58 [18.31]
Sample size	794	629

*Notes: t statistics in parentheses. * denotes statistical significance at the 5% level.*
a Likelihood Ratio test for the significance of the identification restrictions.
Chi squared values at the 5% level are in square brackets.The excluded category are Black Caribbeans, with no qualifications, who arrived in Britain between 1990 and 1994, living in a low ethnic density area in the South of England (excluding London).

[i] See Leslie and Lindley (2001) for a theoretical framework on the acquisition of language skills.

[ii] See Maddala (1983) page 244 for a discussion.

[iii] The Blinder-Oaxaca decomposition method was refined by Neumark (1988) and subsequently developed by Oaxaca and Ransom (1994). However these approaches are not adopted here in an attempt to keep the analysis simple.

[iv] The Blinder-Oaxaca decomposition method involves a comparison of two communities, in this case fluents, F and non-fluents, NF. Here fluents are advantaged relative to non-fluents. The earnings for fluents are given by $Y_F = \alpha_F X_F + u_F$, where Y_F is a vector of the log of earnings, α_F is a vector of coefficients, X_F is a matrix containing the associated characteristics and u_F is a vector containing the residual terms. The latter incorporates unmeasurable characteristic and coefficient effects. A similar earnings function exists for non-fluents. If the mean of the residual terms in these two equations is zero, then the mean of earnings for the two communities is given by $\overline{Y}_Z = \alpha_Z \overline{X}_Z$, where Z = F, NF and the mean pay gap between fluents and non-fluents is therefore $\overline{Y}_F - \overline{Y}_{NF} = \alpha_F \overline{X}_F - \alpha_{NF} \overline{X}_{NF}$.

[v] See Blackaby *et al* (1994) for an application using *the Quarterly Labour Force Survey* from 1992(Q4) to 1995(Q4).

[vi] See Smith and Prior (1996) and Modood *et al* (1997) for a precise description.

[vii] In the FNSEM, data from the 1991 Census was used to divide all electoral wards in England and Wales into three categories according to the percentage of ethnic minorities residing within them. There were (A) whether there were 10 percent or more, (B) whether there were between 0.5 percent and less than 10 percent and (C) whether there were less than 0.5 percent ethnic minorities living in the area. For whites, wards were selected with probability of the number of whites living there. For wards of type A and B an equal number of addresses were sampled in each ward, whilst for wards of type C addresses were sampled in each ward using a constant sampling fraction. At each household found to contain white adults, one of them was selected at random to be the respondent. The procedure is much more complicated for non-whites. In summary for type A and B wards, wards with over 1 percent of Bangladeshis were oversampled by a factor of three, to ensure an adequate sample of South Asians. In type A wards, an interview visited every selected address to find ethnic minority details. In wards of type B, a procedure known as *focused enumeration* was used. This selected every sixth address in the find any ethnic minorities and details were then obtained. In type C wards, the same wards were used as for the white sample. See the appendix in Modood *et al*. (1997) for a detailed discussion.

[viii] Again Smith and Prior (1996) and the appendix in Modood *et al* (1997) provide a detailed description.

[ix] Black African's are not included in the survey.

[x] Dustmann and van Soest (1998a, 1998b) show that self-reported language proficiency measures systematically misclassify language ability and as a result underestimate the importance of language on earnings.

[xi] See Heckman (1979) for a discussion of this now standard procedure.

[xii] This is the procedure adopted by Chiswick and Miller (1995) and Stromback (1986). Miller (1989) also shows that there are little gains from using grouped data techniques rather than this ordinary least squares midpoint procedure.

[xiii] Host country duration is excluded from table (2) since the focus of attention is on the distinction between British and Foreign born non-whites. Also the duration question is not asked to the whites in the sample.

[xiv] There are only 10/180 Black-Caribbean men and 5/249 Black-Caribbean women in this sample that are non-fluent.

[xv] In this sample there are 46 percent of non-fluents who had their interview conducted partially in English. This leaves 27 percent of non-fluents that were interviewed wholly in English and 27 percent that were interviewed wholly in another language.

[xvi] Justification for these exclusion restrictions is based on their exogenous relationship with earnings.

[xvii] The sample is constrained to employees and excludes the self-employed.

[xviii] According to Bell (1997) the 1960's and 1970's saw increases in immigrants from India, East Africa, the Caribbean and Pakistan. However from the 1980's onwards there were large declines in the flows of immigrants coming from India and East Africa and rises in the numbers coming from Ireland and Europe.

[xix] There are 282 non-fluents out of 359 in the sample that have no qualifications