

English language proficiency and earnings in a developing country: the case of South Africa

Daniela Casale and Dorrit Posel
(casaled@ukzn.ac.za) (posel@ukzn.ac.za)

University of KwaZulu-Natal

April 2010

Paper prepared for the 5th IZA/World Bank Conference: Employment and Development

Abstract

In this paper we explore the relationship between English language proficiency and earnings in South Africa, using new data from the first wave of the National Income Dynamics panel survey of 2008. Much of the literature on this topic has studied the impact on earnings of host country language acquisition among minority groups of immigrants to developed countries. In our study we analyse the returns to language skills in a developing country context where the dominant language of business, government and education is that of the former colony, although not more than one percent of the African majority population group speaks English as their home language. Our findings suggest large returns among Africans to reading and writing English very well, and particularly among those who have a tertiary education. We also briefly consider the implications of these results for language and education policy in South Africa in the post-apartheid period.

JEL codes: J24, J31, I21

Keywords: language proficiency, earnings, South Africa, language policy

1. Introduction

Since the late 1970s, a growing body of research has developed on the economics of language. A subset of the literature explores the effects of language on earnings. Most of this work has been conducted in a developed country context and focuses on the effects of destination language proficiency on the earnings of immigrants. Studies have looked particularly at the returns to dominant language fluency among immigrants in the United States (cf. Chiswick 1978; Kossoudji 1988; Chiswick 1991; Chiswick and Miller 1995; Bellante and Kogut 1998; Chiswick and Miller 2002), Germany (Dustmann 1994; Dustmann and van Soest 2001; 2002), Canada, Australia, Israel (Chiswick and Miller 1995; Chiswick 1998; Pendakur and Pendakur 2002) and the United Kingdom (Shields and Wheatley Price 2002). This focus on immigrant language skills forms part of a broader literature concerned with the nature of migrant assimilation into the labour market of the host or destination country.

In a developing country context, language heterogeneity among resident populations derives from a different source. Developing countries typically are not destination countries for immigrants. However, the dominant language in which business and public life is conducted may not be the language spoken by the native or indigenous population. Research which has been undertaken for developing countries explores the relationship between dominant non-indigenous language skills and earnings among indigenous language speakers. Studies look particularly at countries in Latin America, where indigenous language speakers are linguistic minorities (cf. Chiswick et al's 2000 study of Bolivia). There has been relatively little work on developing countries more broadly, however, a gap which Chiswick et al (2000: 350) largely attribute to the scarcity of data on language use.

South Africa offers a particularly interesting opportunity to explore the effects of dominant non-indigenous language skills on earnings. The majority of South Africans speak an indigenous African language as their home language; just over 80 percent of the population are African, of whom less than one percent speak English at home (and an even smaller percentage speak Afrikaans, a derivative of Dutch, at home). However, given the legacy of British and Dutch colonial rule, as well as language policy under apartheid, English and Afrikaans have been the dominant languages of

the economy and government. In 1996, the South Africa government adopted a progressive constitution, giving official status to 11 languages (nine of which are African). However, English in particular remains the dominant language in the public and economic spheres.

Although South Africa has also adopted what is seen to be a very progressive Language-in-Education Policy which encourages schools to teach learners in their home language (Brock-Utne and Holmarsdottir 2004; Probyn 2009), most African schools have rejected a mother tongue policy in favour of maintaining English as the language of instruction. Many African schools start teaching in English from Grade 4 and often from as early as Grade I (Probyn et al 2002). An increasing body of research however suggests that where young African learners have little exposure to English in their daily lives outside the classroom, and where schools are poorly resourced, the early adoption of English as the medium of instruction has been counterproductive, and has adversely affected the acquisition of both English language and African home language skills, as well as the quality of educational attainment (Heugh 1999, Probyn et al 2002; Brock-Utne and Holmarsdottir 2004; Heugh 2007; Probyn 2009).

The returns to English language proficiency among African mother tongue speakers in South Africa have not been investigated empirically because labour force surveys have not collected information on English language ability. The only national empirical study, conducted by Cornwell and Inder (2008), is limited to exploring the relationship between the language most often spoken at home, and labour market outcomes. However, because home language is so highly correlated with race in South Africa, it is difficult to disentangle the effects of race from the effects of language in the earnings estimations.

In this paper, we add to the literature on the returns to dominant language skills in developing countries by exploiting data, collected in a new national household survey, on both home language and English language proficiency in South Africa. We test the relationship between English language proficiency and earnings for African men in wage employment using Ordinary Least Squares estimations although we do consider the effect that possible endogeneity might have on the results. We find very high

returns to English language proficiency, which suggest that large productivity effects and more efficient job search are associated with good English language skills. In contrast, controlling for English language proficiency, we find no independent returns to African language proficiency in the South African labour market.

Very large and convex returns to education are also dampened after controlling for English language proficiency, suggesting that part of the premium to higher education derives from being proficient in English. We investigate this further by exploring the relationship between English language proficiency and earnings at different levels of education. We find particularly large returns to English language proficiency among those who report having completed either a degree or a diploma. Our results are consistent with a strong degree of complementarity between language capital and other forms of human capital. In the South African context, our results may also indicate that English language proficiency acts as a signal to employers about the quality of higher education that individuals have received.

The paper is structured as follows. In the next section, we briefly review the empirical literature which investigates the relationship between dominant language proficiency and earnings and we elaborate on the particular South African context. In Section 3, we discuss the data that we analyse in the study, how we define language proficiency, and the method of estimation. In Section 4, we present and discuss the findings of the study, and in the final section, we conclude and consider briefly the implications of the results for language policy in education in South Africa.

2. Background

Research on the returns to dominant language skills is rooted in the human capital earnings literature. Language proficiency is an important part of human capital (Chiswick and Miller 1995) which is expected to increase earnings by directly raising productivity or by acting as a signal to the employer of the individual's potential productivity and ability. Dominant language proficiency also lowers the transaction and information costs of job search for the individual, and it increases the ability to

negotiate the terms of employment and the possibility of accessing better paid jobs (Dustmann 1994; Shields and Wheatley Price 2002).

Much of the work on the returns to language capital has explored how destination language ability affects the earnings of immigrants in developed countries, where language ability is measured typically in terms of speaking fluency. Results from these studies generally identify a positive and significant return to dominant language fluency, with the size of the premium ranging from about 5 to 20 percent. In an international comparison of fluency in the destination language among immigrants in the 1980s, for example, Chiswick and Miller (1995) estimate an earnings premium of between 5.3 and 9.3 percent in Australia, 16.9 percent in the United States, 12.2 percent in Canada and 11 percent in Israel. Dustmann and van Soest (2002) identify a premium of approximately 5 to 15 percent, depending on the method of estimation, to dominant language fluency among immigrants in Germany from 1984 to 1993.¹

A related question that has been explored in this literature concerns the relationship between language skills and other forms of human capital. Among immigrants with previously acquired formal education, dominant language fluency would be expected to increase the “international transferability” of this education (Chiswick and Miller 2002: 40). In their study of immigrants in the United States, Chiswick and Miller (2002) find far higher returns to an additional year of schooling among immigrants fluent in English than among those not fluent (6.6 percent compared to 1.0 percent), results which suggest a strong degree of complementarity between language and human capital (see also Chiswick and Miller 1995).

There has been far less research which investigates the returns to dominant language skills in developing countries, partly because of data constraints on language usage

¹ Other work on the relationship between language ability and earnings in developed countries investigates the implications for minority language mother-tongue speakers in multilingual or bilingual countries. Grin and Sfreddo (1998), for example, estimate the effects of speaking Italian as a first language in Switzerland, where Italian is an official but minority language (spoken by less than 10 percent of residents), and where Italian speakers are concentrated in particular areas of Switzerland. A body of research also investigates the returns to bilingualism in Canada (see Pendakur and Pendakur 2002 for a brief review).

(Chiswick et al 2000). Developing countries also are not typically the destination countries of immigrants. Given the legacy of conquest and colonial rule, however, in many developing countries the dominant language is not the home language or mother tongue of the indigenous population. In one of the few studies which investigate this linguistic heterogeneity, Chiswick et al (2000) explore earnings differences in Bolivia, among monolingual Spanish speakers, indigenous language only speakers, and individuals who speak both an indigenous language and Spanish (and who are identified as “bilingual” speakers). They find that monolingual Spanish speakers earn about 25 percent more than bilingual speakers, and surprisingly, that bilingual speakers have no earnings advantage over indigenous language only speakers. However, because respondents were not asked to identify how well they could speak Spanish, there may be large variation in Spanish language proficiency among those individuals identified as bilingual speakers.

South Africa is an example of a developing country in which the majority of the population speaks an indigenous language which is not the language of business or politics. Prior to 1996, as a consequence first of colonialism and then of White-dominated rule, English and Afrikaans were the country’s official languages. The end of apartheid, and the adoption of the new constitution in 1996, brought “language democratization” (Kamwangamalu 2002: 120), and nine African languages, in addition to English and Afrikaans, were granted equal and official status. The new constitution also requires of the state that it take “practical and positive measures to elevate the status and advance the use of these (African) languages” (Constitution of the Republic of South Africa 1996, Section 6(1)). The reality in post-apartheid South Africa, however, is that English remains the dominant language, and although only ten percent of South Africans speak English as their home language, the language has “gained in power ... at the cost of Afrikaans and African languages” (Probyn 2009: 126).

Under apartheid, all formal education was racially segregated, and resources were disproportionately allocated to White schools. Although English and Afrikaans were South Africa’s official languages, apartheid education policy required that Africans were educated in their mother tongue in primary school, after which instruction would occur in either English or Afrikaans. Following the Bantu Education Act of 1953,

mother-tongue education for African language speakers was extended from Grade 4 to Grade 8, an extension which was “designed to further separate development and prevent African language speaking students from developing ambitions outside their own communities” (Heugh 1999: 302). Attempts to then enforce Afrikaans as the medium of instruction for half the curriculum in secondary schools in 1975 sparked the student-led uprisings in Soweto in 1976. Resistance to compulsory learning in Afrikaans became coupled with the rejection of mother-tongue education in African schools, and in the face of escalating opposition, the Education and Training Act was passed in 1979. The act reduced mother tongue instruction among African language speakers to four years of primary school, and thereafter schools could choose whether the medium of instruction would be English or Afrikaans, with most schools choosing English.

With the transition to democracy, the post-apartheid government adopted what is seen to be a very progressive Language-in-Education Policy (LiEP), which devolves the decision on the language of learning and teaching to the individual schools (through the School Governing Bodies) (Probyn et al 2002; Brock-Utne and Holmarsdottir 2004). The policy promotes (but does not mandate) the use of African languages alongside English, encouraging schools to maintain the learners’ home languages at the same time as they learn an additional language. The policy is based on the notion of additive bilingualism, developed in theories of second language acquisition, that “a thorough conceptual grounding in the learners’ home languages provid(es) the basis for later learning through an additional language” (Heugh 1999: 127).² Most schools, however, have not conformed to the proposals of LiEP, and have maintained the use of English as the language of learning and teaching from at least Grade 4.

African mother-tongue education in South Africa has been tainted by its association with the repressive policies of apartheid, and many schools have opted for the early introduction of English as the medium of instruction because English is widely perceived as “the language of upward mobility and access” (Probyn 2009: 126).

² Heugh (1999) notes that over the past 90 years, many commissions of enquiry into language-in-education policy in African countries have attested to the importance of mother-tongue (indigenous) languages as the initial language of learning.

However, this has had unintended consequences. Young African learners, particularly those from poorer communities who attend historically disadvantaged (African) schools, do not have sufficient time in the classroom, and enough exposure to English outside the classroom, to acquire a threshold level of English language proficiency which allows them to engage effectively with learning in English. Heugh (1999: 304) explains it thus:

“African language speaking students have, for the last two decades, been obliged to undergo too rapid and too early a change from the use of the mother tongue as the medium of learning for effective or successful transfer to occur”.³

By retaining, and sometimes even hastening the early introduction of English as a medium of instruction, many studies suggest that language policy in schools has been counterproductive (cf. Heugh 2000; Probyn et al 2002; Brock-Utne and Holmarsdottir 2004; Heugh 2007; Probyn 2009). It has compromised the acquisition of language skills in both the learner’s African mother tongue and English and has contributed to a widening language proficiency and educational gap between the historically disadvantaged schools, located in African townships and rural areas, and the historically advantaged (formerly White and now desegregated) schools, located mainly in urban centres.⁴ Language skills among Africans therefore have increasingly become a marker of the quality of education attained.

There have been no national studies which investigate levels of English language proficiency in South Africa. Moreover, although English language skills are widely believed to have significant economic currency in the labour market, no research has

³ The English language proficiency of teachers who are themselves not English mother-tongue speakers may also be “inadequate for effective teaching and learning to occur through English” (Heugh 1999: 309). Teachers resort to code-mixing (switching languages within sentences) and code-switching (switching languages between sentences) in the classroom (Probyn et al 2002; Brock-Utne and Holmarsdottir 2004), although school-leavers are required to write their final (matriculation) examinations in English.

⁴ Heugh (1999: 303) points out that ironically, during the years of Bantu education when African mother-tongue education was extended from Grade 4 to Grade 8, African matriculation rates improved. The rejection of mother tongue policy therefore misdiagnosed the source of the problem with “Bantu education”, which was rather the poor quality of the curriculum, compulsory instruction in secondary school in both English and Afrikaans, and a significant lack of resources for African schools.

empirically estimated the earnings advantage to English language proficiency or explored the relationship between English language proficiency and the returns to formal education. The study by Cornwell and Inder (2008) is the first to investigate the link empirically between English language and earnings using a Mincerian earnings function and nationally representative survey data.⁵ However, the only information which these authors have available for their analysis is the language which the individual speaks most often at home. No information is collected on whether an individual is also proficient in another language.

The study finds an earnings premium to all English home language speakers of about 17.7 percent, and further that the premium to being White falls to almost half the value estimated in a regression without language controls (Cornwell and Inder 2008). With less than one percent (0.32 percent)⁶ of the sample of Africans reported as English home language speakers in their data, however, a high degree of multicollinearity between race and home language confounds these estimates. Furthermore, the small number of Africans who report English as their home language is likely to represent a select group with a privileged upbringing and access to education (a selection bias for which their study cannot control).

Without information on English language proficiency among African home language speakers, it therefore has not been possible to disentangle the returns to English in the South African labour market from other effects. Our study seeks to fill this research gap by analysing data from the first nationally representative household survey to collect information on English language and home language proficiency in South Africa. We estimate the size of the earnings advantage to English language proficiency and we explore the relationship between English language proficiency and earnings at varying levels of education among African men in the South African labour market.

⁵ Cornwell and Inder (2008) estimate earnings for a pooled sample of individuals in wage and self-employment drawn from three rounds of the October Household Survey (1996 to 1998).

⁶ Although Cornwell and Inder (2008: 491) report that 3.2 percent of Africans in their sample identify English as their home language, this figure seems to have been misreported. Table 1 of their study (2008: 501) shows that among Africans, 404 of a sample of 124 262 (or 0.32 percent) are identified as English home language speakers.

3. Data and methods

The data used for this study are drawn from the baseline wave of the National Income Dynamics Survey (NIDS) which was conducted in 2008 and released in July 2009. NIDS is the first national household panel survey in South Africa, designed to track approximately 28 000 individuals in just over 7 300 households at two-year intervals. Although funded by the South African National Treasury, the survey was conducted by the Southern Africa Labour and Development Research Unit (SALDRU) at the University of Cape Town's School of Economics, and involved extensive consultation with academics, policy-makers and other potential users. The questionnaires cover a wide array of topics including births, deaths, migration, labour market participation, education, health, and income and spending. The survey is unique in South Africa not only because of the panel design, but also because these key areas of socio-economic enquiry are explored in one survey, whereas previous national household surveys have tended to focus on a particular area such as labour market participation, health or income and spending.

The NIDS data are particularly useful for our purposes as information is collected on both labour market outcomes and language proficiency. In the South African Population Census and in other national household surveys, individuals are only asked what language they speak most at home. NIDS is the first national survey which also captures information on whether individuals are proficient in their home language and in addition, whether they are proficient in English.

3.1 Home language and English language proficiency

The NIDS survey asked individuals how well they could read and write English and their home language in two separate sets of questions, with response options on a four-point scale of “very well”, “fair”, “not well” and “not at all” (questions H36 to H39 in the adult questionnaire). Following other work in this area (Chiswick and Miller 1995; Rendon 2007), we chose a strict definition of proficiency to minimise concerns about the over-reporting of language ability (Dustmann and van Soest 2001;

Deumert and Mabandla 2006). We define proficiency as the self-reported ability to read and write “very well” in the specified language.

In the South African data, there is a very large overlap between reading and writing English “very well”. Over 96 percent of those who report reading very well in English also report writing very well in English (with very similar results among Africans specifically). The correlations in the other categories of “fair”, “not well” and “not at all” are also all over 90 percent, suggesting that respondents were consistent in the way they responded to these questions. A similar pattern emerges for reading and writing skills in the home language⁷, for the full sample of adults and for Africans specifically.

In the international literature, proficiency has typically been defined in terms of self-assessed ability to speak a language, as most large-scale surveys do not collect information on reading or writing skills (Chiswick and Miller 1995; Bellante and Kogust 1998; Chiswick and Miller 2002). In the few studies that have data on more than one of these language skills, the evidence generally suggests that reading and writing proficiency are more important (larger and more significant) determinants of earnings than speaking fluency (see for example, Chiswick 1991 and Dustmann 1994). Reading and writing skills are likely to provide a better measure of the individual’s knowledge of a language than speaking skills, and in addition, many higher-paying jobs will require the individual to be able to read and write to perform the functions of the job efficiently.

The data in Table 1 help to explain why using an individual’s home language to approximate English language proficiency in South Africa is limiting. Among the

⁷ There were some anomalous results in the reporting of proficiency in home language. Eighty-two individuals, the majority Indian, report English as the language they usually speak at home and being proficient in English, yet they report not being proficient in their home language. This is likely to reflect differences in the interpretation of the ‘language usually spoken at home’ (which is asked at the beginning of the adult questionnaire) and ‘home language’ in the question on proficiency (which is asked towards the end of the questionnaire) among this group. It suggests that what may have been interpreted as home language in the latter is the individual’s language of ‘origin’ or ‘ancestral’ language rather. These responses were recoded to proficient in home language for consistency.

sample of working-age Africans (aged 25 to 65 years old), only about one percent report English (and even less Afrikaans) as their home language, while the vast majority (97 percent) report one of the official African languages as the language usually spoken at home. Among Africans of this age group, the few who report English as their home language are likely to be a very select group of individuals, who have perhaps schooled at one of the country's top private English-medium institutions or have grown up or spent a substantial amount of time living in English-speaking countries during the apartheid years. Similarly, among Coloureds, Indians and Whites, only a very small minority speak one of the African languages at home (three, two and one percent respectively). The majority of Coloureds and Whites (96 and 98 percent respectively) report either English or Afrikaans as their home language, while most Indians (91 percent) speak English at home.⁸

Although there is a very high degree of home language segregation across the race groups, the incidence of bilingualism or multilingualism in South Africa is also high. This is illustrated by comparing the first two columns of Table 2. Only one percent of working-age Africans report English as the language they usually speak at home, but about 35 percent are identified in the NIDS data as being proficient in English. There are also large differences in the percentages reporting English as a home language and English language proficiency among Coloureds and Whites, as the majority in these two groups report speaking Afrikaans most often at home.

The data in Table 2 (third column) highlight a large racial variation in home language skills. For example, only 58 percent of Africans report being able to read and to write very well in their home language compared to 95 percent of Whites. This variation reflects decades of uneven educational opportunity in South Africa. The NIDS data indicate that among a sample of 25 to 65 year old adults, almost 40 percent of Africans either have no schooling or only some primary schooling, compared to less than one percent of Whites. Furthermore, as we noted earlier, many studies have suggested that home language skills among Africans have also been compromised by

⁸ The Labour Force Survey of September 2007, conducted about six months before the start of the NIDS survey, also asks a question on the language spoken most at home. The data produce a similar language distribution by race to that found using the NIDS data (based on own calculations).

the early adoption of English as the language of instruction in historically disadvantaged African schools.

Table 1. Home language and race in South Africa (percentages)

	African	Coloured	Indian	White	Total
IsiNdebele	1.47 (0.211)	0.32 (0.303)	0.00 (0)	0.00 (0)	1.15 (0.164)
IsiXhosa	21.94 (0.694)	1.23 (0.498)	1.312 (1.159)	0.00 (0)	16.93 (0.559)
IsiZulu	28.93 (0.766)	1.33 (0.836)	0.93 (0.665)	0.00 (0)	22.27 (0.630)
Sepedi	12.90 (0.616)	0.00 (0)	0.00 (0)	0.07 (0.071)	9.87 (0.484)
Sesotho	12.80 (0.566)	0.16 (0.104)	0.00 (0)	0.04 (0.038)	9.80 (0.444)
Setswana	10.28 (0.486)	0.42 (0.194)	0.00 (0)	0.99 (0.982)	8.01 (0.393)
SiSwati	2.75 (0.258)	0.00 (0)	0.00 (0)	0.00 (0)	2.10 (0.198)
Tshivenda	1.81 (0.217)	0.00 (0)	0.00 (0)	0.03 (0.028)	1.38 (0.166)
Xitsonga	4.03 (0.370)	0.00 (0)	0.00 (0)	0.00 (0)	3.08 (0.285)
Afrikaans	0.96 (0.181)	70.45 (2.820)	5.29 (3.889)	57.22 (3.307)	13.94 (0.615)
English	1.28 (0.298)	25.77 (2.795)	91.0 (4.255)	40.71 (3.277)	10.63 (0.671)
Other	0.87 (0.256)	0.32 (0.317)	1.48 (1.467)	0.95 (0.703)	0.84 (0.218)
Sample size	100 6 916	100 1 511	100 155	100 621	100 9 203

Source: NIDS 2008.

Notes: The data are weighted. Standard errors are in parentheses. The sample includes all adults aged 25 to 65 years.

Table 2. English home language and language proficiency (percentages)

	English as home language	English language proficiency	Home language proficiency	English and home language proficiency
African	1.28 (0.298)	35.31 (0.843)	58.01 (0.832)	33.86 (0.837)
Coloured	25.77 (2.795)	46.67 (2.648)	67.73 (2.367)	45.80 (2.659)
Indian	91.00 (4.255)	87.47 (3.183)	86.28 (3.423)	85.74 (3.461)
White	40.71 (3.277)	83.19 (2.389)	95.34 (1.216)	82.95 (2.396)

Source: NIDS 2008.

Notes: The data are weighted. The sample includes all adults aged 25 to 65 years. Language proficiency is defined as the self-reported ability to read and speak the language “very well.” Home language is the language the individual usually speaks at home.

3.2 Methods

In this study, we test the relationship between proficiency in English and earnings in the South African labour market. We estimate a standard Mincerian earnings equation for a sample of men⁹ aged 25 to 65 years who have wage employment:

$$\ln(W_i) = \alpha + \gamma E_i + \beta X_i + \varepsilon_i \quad (1)$$

The dependent variable, W_i , is the log of individual hourly wages for individual i ; E_i represents an indicator equal to one if the individual is proficient in English and zero otherwise; X_i is a vector of observable characteristics including age, education, marital status, area type (urban or rural) and province of residence; and ε_i is the error term. To compare the returns to English language proficiency and African language proficiency, in some of the specifications we include in X_i a dummy variable indicating whether the individual can read and write very well in one of the nine official African languages. The means and standard errors of the explanatory variables are provided in the Appendix (Table A1).

We also investigate the complementarity argument in the language and earnings literature that there are greater benefits to other forms of human capital, such as formal education, among those with better English language skills. In a separate specification we include a set of interaction terms between English language proficiency and levels of education

$$\ln(W_i) = \alpha + \gamma E_i + \gamma E_i * Educ1_i + \gamma E_i * Educ2_i + \gamma E_i * Educ3_i + \beta X_i + \varepsilon_i \quad (2)$$

where $Educ1_i$ represents whether individual i has only some primary or secondary schooling; $Educ2_i$ represents a completed secondary education (i.e. that the individual

⁹ With some exceptions (for example, Dustmann 1994; Chiswick et al 2000; Dustmann and van Soest 2002), the literature on this subject restricts the analysis to males to avoid the problem of the non-random participation of women in the labour market. We follow a similar practice here.

has passed the final matriculation exams) and $Educ3_i$ indicates whether the individual has some form of tertiary education, i.e. either a diploma or a degree.

There are concerns in the empirical literature regarding the validity of the estimates obtained from Ordinary Least Squares (OLS) estimation of earnings on language proficiency for two main reasons - namely, that there may be measurement error in the language proficiency variable and that language proficiency itself may be endogenous to earnings.

In the first case, measurement error in the data will lead to an underestimation of the effect of language proficiency on earnings. Using panel data for Germany, Dustmann and van Soest (2001; 2002) estimate the impact of measurement error in self-reported speaking fluency, and find that people tend to over-report rather than under-report fluency. In their qualitative study of language and communication networks in four low-income areas in Cape Town, South Africa, Deumert and Mabandla (2006) also find over-reporting of English proficiency among the respondents. We do not have the sort of panel data that Dustmann and van Soest (2001; 2002) were able to use to correct for measurement error. However we have sought to minimise this problem by using a stringent definition of English language proficiency which requires individuals to both read *and* write English “very well” (Chiswick and Miller 1995; Rendon 2007).

In contrast, the problem of endogeneity of language skills in an earnings regression is likely to produce an upward bias in the coefficient on English language proficiency. This would be the case if there is unobserved heterogeneity that is positively correlated with both second-language acquisition skills and earnings. The typical example used in the literature is the effect of unobserved ability on language acquisition (Dustmann and van Soest 2002; Shields and Wheatley Price 2002). Pendakur and Pendakur (2002: 150) express this as follows: “People who have the ability to speak more than one language may also have other characteristics that allow them access to higher incomes, such as the ability to learn quickly and effectively”. The endogeneity problem may also derive from the reverse dependence of language on expected earnings (Chiswick et al 1995; Chiswick 1998). For example, those who stand to benefit the most from language proficiency, through better employment

opportunities or higher earnings, may also invest the most in acquiring these language skills.

This latter cause of endogeneity is likely to be less of a problem in the South African context than in studies on immigrants and host country language skills. Immigrants often arrive with few or no host country language skills and need to make decisions about whether and how to acquire these skills. In South Africa, English language proficiency will be influenced largely by the language of instruction and quality of education during schooling, which predates labour market entry. This is especially the case for reading and writing skills which are harder to acquire informally on the job or through casual communication with neighbours and co-workers.¹⁰

However, the existence of unobservable characteristics that affect both English language acquisition and earnings is a concern in this study. In the international empirical literature a variety of instruments have been used to address this problem, but these have been mostly in the context of immigrants' acquisition of host country language skills and are therefore not suitable to the South African context. For example, Chiswick et al (1995) and Chiswick (1998) in their studies of immigrants to Australia and Israel respectively, instrument with the number and age of children in the household, whether the individual was married overseas and minority language concentration in the area. Shields and Wheatley Price (2002) use a set of variables reflecting whether the interview was conducted wholly or partially in English or another language, whether the immigrant was married to a UK born spouse and the number of dependent children in the household. Dustmann and van Soest (2001; 2002) use father's education¹¹ in a panel data model on immigrants in Germany.

¹⁰ Dustmann (1994) finds evidence that in contrast to speaking proficiency, writing proficiency is not affected by labour market attachment as this form of language knowledge requires more systematic learning, which will often occur during schooling.

¹¹ This variable has been questioned as a valid instrument in an earnings equation as father's education may have a direct effect on earnings through networking opportunities. However, Dustmann and van Soest (2001; 2002) argue that this is less of a problem among their sample of first-generation immigrants in Germany as networking by parents to improve labour market outcomes for their children is unlikely to occur.

We were not able to find any suitable instruments in the NIDS data that would be correlated with English language proficiency but not with the error term in the earnings equation. Our estimations therefore may overstate a positive relationship between English language proficiency and earnings. The instruments which we considered and the implications for our findings are discussed further in the next section.

4. Estimations

The results for the first set of estimations are presented in Table 3. The first column reports the OLS estimates from a basic earnings regression on the log of hourly wages excluding the dummy variable for English language proficiency. The coefficients all have the hypothesised signs, and the results are generally comparable to those found in other studies on the determinants of earnings in South Africa among African men using a similar specification (Hofmeyr and Lucas 2001; Casale and Posel, 2009; 2010). In particular, there are very high returns to completed secondary (matric) and post-secondary education compared to the reference category of no schooling. Individuals with a matric earn on average 120 percent more than those without any schooling and about 330 percent more if they have some post-secondary qualification. These highly convex returns to education among Africans have been explained in terms of a “rationed supply hypothesis” (Mwabu and Schultz 2000): apartheid education policies limited the supply of more highly skilled African labour while the demand for this labour has increased dramatically in the post-apartheid period.

The second specification, which includes the variable representing English language proficiency, suggests a large and significant premium to reading and writing very well in English among African males. Those who are proficient in English have an earnings advantage of almost 55 percent over those who are not proficient in English. This may be because labour market participants who are English language proficient are more capable of finding and securing work that would provide them with higher returns. Furthermore, proficiency in English is likely to increase the productivity of the worker on the job because he is better able to perform the tasks of the job, and

communication costs across workers, and between workers and management, are lower.¹²

Table 3. Earnings estimations and English language proficiency, African males, 25 to 65 years

	I	II	III	IV
Age	0.095*** (0.033)	0.089*** (0.031)	0.089*** (0.032)	0.087*** (0.031)
Age ² /100	-0.097*** (0.037)	-0.089** (0.036)	-0.090*** (0.036)	-0.087** (0.036)
Incomplete secondary	0.246* (0.134)	0.176 (0.134)	0.167 (0.134)	0.153 (0.134)
Secondary/matric	0.792*** (0.171)	0.492*** (0.182)	0.648*** (0.169)	0.460*** (0.176)
Post-secondary	1.456*** (0.188)	1.154*** (0.192)	1.330*** (0.185)	1.129*** (0.186)
Married	0.195** (0.099)	0.176* (0.094)	0.196** (0.099)	0.178* (0.094)
Urban	0.247** (0.102)	0.193* (0.101)	0.226** (0.103)	0.189* (0.102)
English proficiency	--	0.437*** (0.095)	--	0.414*** (0.108)
African language proficiency	--	--	0.191** (0.086)	0.063 (0.094)
Constant	-0.495 (0.732)	-0.321 (0.719)	-0.348 (0.707)	-0.277 (0.702)
F stat	13.41	13.12	12.67	12.34
Prob > F	0.00	0.00	0.00	0.00
R ²	0.292	0.316	0.298	0.317
Sample size	1085	1084	1085	1084

Source: NIDS 2008

Notes: The data are weighted. Standard errors are in parentheses. *** Significant at the 99 percent confidence level. ** Significant at the 95 percent confidence level. * Significant at the 90 percent confidence level. The dependent variable is the log of hourly wages. The regressions also include controls for the nine provinces in South Africa. The omitted categories are no schooling, not married, rural residence, not proficient in English and not proficient in an African language.

¹² Following the practice in the international literature (McManus et al 1983; Chiswick 1991; Chiswick and Miller 1995; Dustmann 1994; Dustmann and van Soest 2001), we do not control for job characteristics in the regressions because we want to estimate the returns to English language proficiency including the benefits of language capital in gaining access to particular types of employment. There is very little information collected on job characteristics in the NIDS survey across all types of wage employment. However, when we include a set of dummy variables representing the nine main occupational categories (based on the one-digit International Standard Classification of Occupation codes), as expected the coefficient on English language proficiency falls, but only from 0.437 to 0.366 and it remains highly significant at the one percent level.

The impact of the other variables on earnings remains largely unchanged when English language proficiency is included in the regression, except for the education variables. In comparison to no schooling, the returns to incomplete secondary education are no longer significant, while the returns to completed secondary fall by almost 60 percentage points and those to post-secondary education decline by 110 percentage points. The substantial reduction in the returns to education, and higher education in particular, suggests that part of the earnings premium to education in South Africa derives from English language proficiency. This may be because in South Africa, in addition to the job-search and productivity benefits mentioned above, English language proficiency acts as a signal to employers of the quality of education that the worker has received. Given that there are large concerns about the quality of schooling available to Africans in both pre- and post-apartheid South Africa (particularly in rural and township areas, and in the former homeland areas of South Africa), English language proficiency is likely also to reflect whether the individual was educated in a formerly advantaged school or university previously reserved for Whites, a private institution, or one of the better formerly disadvantaged schools or universities.

The last two columns of Table 3 present the specifications that include an indicator for proficiency in an African language. As we would expect given the dominance of English in both the business and government sectors in South Africa, being able to read and write very well in the mother tongue of the vast majority of the sample does not offer the same economic value as being able to read and write very well in English. When included on its own in the regression, African language proficiency has a coefficient of 0.191 (compared to the English language coefficient in regression II of 0.437). However, when both the English and African language proficiency variables are included in the regression, the premium falls to 0.063 and is no longer significant. In contrast, the premium to being proficient in English falls only slightly to 0.414, and it remains highly significant. In our sample, approximately 60 percent of all men who are African language proficient are also English-language proficient. However, 98 percent of men who are English language proficient are also proficient in an African language. The positive coefficient on the African language proficiency variable in regression III is likely to reflect this relationship between African language and English language proficiency. The inclusion of the English language proficiency

variable in regression IV, however, suggests that for men who do not have good English language skills, there are no independent economic advantages to being proficient in an African language.

Table 4 presents the estimates from the regression which tests the relationship between English language proficiency and earnings at varying levels of education. While the premium to those proficient in English remains high, at 36 percent, compared to those who are not proficient, the coefficients on the interaction terms suggest that there is a very large additional premium to employed African men who are both English language proficient and who have completed either a diploma or a degree. Being able to read and to write very well in English and having completed some form of post-secondary qualification offers an additional premium of almost 97 percent. It is also interesting to note that the coefficients on the education variables representing a completed secondary or post-secondary education are now no longer significantly different from each other. This implies that the benefits to a tertiary qualification accrue only to those who also report being English language proficient.

These results are consistent with the complementarity argument that individuals are able to make better use of their investment in other forms of human capital such as education when they are proficient in the dominant language. Again, they may also reflect the peculiarity of the South African education system: English language proficiency may be used to “screen” the quality of post-secondary education. In other words, the value of a diploma or degree may be lower in the workplace for those who are not proficient in English because English proficiency provides a strong signal to employers of the quality of that qualification.

Table 4. Earnings and English language proficiency among African males, including interaction terms for education

	OLS on ln(hourly wage)
Age	0.089*** (0.031)
Age ² /100	-0.90*** (0.036)
Incomplete secondary	0.179 (0.137)
Secondary/matric	0.693*** (0.210)
Post-secondary	0.709*** (0.233)
Married	0.179* (0.096)
Urban	0.194** (0.099)
English proficiency	0.309* (0.169)
Incomplete secondary*English proficiency	0.090 (0.226)
Secondary*English proficiency	-0.154 (0.244)
Post-secondary*English proficiency	0.676** (0.285)
Constant	-0.177 (0.727)
F stat	24.80
Prob > F	0.00
R ²	0.325
Sample size	1084

Source: NIDS 2008

Notes: The data are weighted. Standard errors are in parentheses. *** Significant at the 99 percent confidence level. ** Significant at the 95 percent confidence level. * Significant at the 90 percent confidence level.

A concern with these OLS results is that there may be unobservable characteristics, such as ability, that are correlated with both language proficiency and earnings. If this is the case, our estimates of the premium to English language proficiency will be overstated. To correct for this bias, studies in this area have used a variety of instrumental variables which, as we explained earlier, are either not available to us or are not applicable in the South Africa context. We did test two possible exclusion restrictions, but both proved problematic. First, we included mothers' and fathers' education as exclusion restrictions in the model predicting proficiency (see, for example, Dustmann and van Soest 2001; 2002). However, neither of these variables (when included together or separately) had a significant effect on the probability of

being proficient in English. These variables have also been criticised as valid instruments because they themselves may be determinants of earnings.¹³

We also considered using African language proficiency as an exclusion restriction given that African language proficiency and English language proficiency are highly correlated in our data, and in light of the substantial evidence in the language and education literature that a good grounding in an individual's home language is a prerequisite for proficiency in a second language.¹⁴ Furthermore, as the results in Table 3 showed, controlling for the other correlates of earnings, proficiency in the mother tongue of the majority of our sample of African men seems to have no independent effect on earnings. However, while African language proficiency partly satisfies the requirements of a good instrument in that it is highly correlated with the endogenous variable and does not seem to belong in the earnings equation, it may itself be correlated with the error term in the earnings equation. This would be the case if we believe that ability or motivation, for instance, are also correlated with home language skills. While there is a large body of linguistic literature that supports the view that "linguistic competence" (Chomsky 1965) in an individual's mother tongue or home language is not determined by ability,¹⁵ this argument does not

¹³ We also tried including mothers' and fathers' education variables directly in the earnings regressions (with and without English language proficiency) but they did not have any significant effect on earnings and the magnitude and significance of the coefficient on English proficiency were largely unchanged. It is possible that in the South African context, parents' education variables are not good predictors of English language proficiency or earnings because of the lack of variation in these variables. The parents of the individuals in our sample of African men aged 20 to 65 years old would have been schooled during a period when the apartheid education system offered very poor education, and limited options for further schooling, to Africans. For example, just over 90 percent of the mothers and fathers of men in our regression sample did not complete their secondary schooling.

¹⁴ In one of the few studies that has information on home language writing skills as well as host language fluency among immigrants, Dustmann (1994) finds that home language literacy has a very strong effect on foreign language skills (although he does not make any correction for endogeneity in this study).

¹⁵ Chomsky (1965) argues that humans are genetically predisposed to acquire language, meaning that every (healthy) individual has "linguistic competence", or the knowledge of the grammar of his/her language.

extend to the acquisition of further reading and writing skills in the individual's home language.¹⁶

5. Conclusion

In this paper we have explored the returns to English language proficiency among African men in the South African labour market. Almost all the men in our sample are African language speakers, about 40 percent of whom are also identified as being English language proficient, defined here as the self-reported ability to both read and write English very well. We find a very high premium to English language proficiency of over 50 percent among this group of individuals, and in addition that the inclusion of controls for English language proficiency in the earnings regression lowers the high returns to education in South Africa. Our estimations suggest that among African men who are not English language proficient, there is no significant difference in the returns to completed secondary education or post-secondary education.

Further analysis showed that the benefits to being proficient in English accrue particularly to those with a post-secondary education. In the OLS estimations, African men with post-secondary education are estimated to earn approximately 97 percent more if they are also English language proficient. These results are consistent with the argument that there is complementarity between these two types of human capital. They also suggest that, in the South African context, English language skills may be used by employers to screen the quality of higher education among the work force.

In contrast, there is little evidence to suggest that language skills in an individual's African home language are independently rewarded in the South African labour market. In an earnings estimation that excludes a control for English language skills, we find a positive and significant relationship between African language proficiency

¹⁶ Attempts to correct for endogeneity using African language proficiency as the identifying restriction resulted in a fall in the English language coefficient from 0.437 to 0.389, although it remained significant at the five percent level.

and earnings because African language skills are highly correlated with English language skills. In our sample, approximately 60 percent of African men who are identified as proficient in an African language are also proficient in English. Among those who are English language proficient, 98 percent are also proficient in an African language. When both African language and English language proficiency are included in the earnings estimation, being proficient in an African language no longer has an independent effect on earnings while the coefficient for English language proficiency remains large and significant.

Endogeneity in language proficiency will cause our estimates to be upwardly biased. However, we do not have suitable instruments to control for the possibility that the acquisition of language skills, in both an individual's home language and in a second language, may be influenced by unobservable characteristics that also affect earnings. In the context of a labour market in which English is the dominant language of business, however, it seems unlikely that the large return to English language proficiency is explained only by unobservable characteristics, particularly when the estimation controls also for African language proficiency.

Our findings have implications for the debate over language policy in education in South Africa. Higher average earnings among the English language proficient would explain why English is perceived as the language of 'success' and why parents want their children to study in English. Notwithstanding a progressive Language-in-Education Policy in South Africa, most historically disadvantaged African schools have rejected mother-tongue education in favour of English as the language of instruction, sometimes from as early as Grade 1. Research on language policy in education in South Africa, however, suggests that these efforts have been misguided, and that without adequate language skills in their mother tongue, learners are not able to acquire strong second language skills or to learn effectively in a second language. The strong correlation which we find between African language and English language proficiency would be consistent with these arguments of additive bilingualism.

Appendix: Descriptive statistics

Table A1. Characteristics of African men (25 to 65 years) with wage employment.

	Mean	Standard error
Age	37.87	0.4151
No schooling	0.077	0.0105
Incomplete secondary	0.545	0.0224
Complete secondary/matric	0.210	0.0190
Post-secondary	0.169	0.0180
Married	0.436	0.0221
Urban	0.692	0.0191
English proficiency	0.421	0.0223
African language proficiency	0.607	0.0220
Sample size	1 084	

Source: NIDS 2008.

References

- Casale, D. and Posel, D. (2009) "Unions and the gender wage gap in South Africa." Economic Research Southern Africa, Working Paper No. 113.
- Casale, D. and Posel, D. (2010) "The male marital earnings premium in the context of bride wealth payments: Evidence from South Africa." *Economic Development and Cultural Change* 58(2): 211-230.
- Bellante, D. and Kogut, C. (1998) "Language ability, US labor market experience and the earnings of immigrants." *International Journal of Manpower* 19(3): 319-330.
- Brock-Utne, B. and Holmarsdottir, H.B. (2004) "Language policies and practices in Tanzania and South Africa: problems and challenges." *International Journal of Educational Development* 24: 67 – 83.
- Chiswick, B.R. (1978) "The effect of Americanization on the earnings of foreign-born men." *Journal of Political Economy* 86: 89 – 921.
- Chiswick, B.R. (1991) "Speaking, reading and earnings among low-skilled immigrants." *Journal of Labor Economics* 9(2): 149-170.
- Chiswick, B.R. (1998) "Hebrew language usage: Determinants and effects on earnings among immigrants in Israel." *Journal of Population Economics* 11 (2): 253-271.
- Chiswick, B.R. and Miller, P.W. (1995) "The endogeneity between language and earnings: International analyses." *Journal of Labor Economics* 13(2): 246-288.
- Chiswick, B.R and Miller, P.W. (2002) "Immigrant earnings: Language skills, linguistic concentrations and the business cycle." *Journal of Population Economics* 15(1): 31 - 57

Chiswick, B. R., Patrinos, H.A. and Hurst, M.E. (2000) “Indigenous Language Skills and the Labor Market in a Developing Economy: Bolivia.” *Economic Development and Cultural Change* 48(2): 349-367.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

Constitution of the Republic of South Africa (1996) Act 108 of 1996.

Cornwell, K. and Inder, B. (2008) “Language and labour markets in South Africa.” *Journal of African Economies* 17(3): 490-525.

Deumert, A. and Mabandla, N. (2006) “Language, communication networks and economic participation: Towards an inclusive economy.” Paper presented at the DPRU conference on “Accelerated and Shared Growth in South Africa: Determinants, Constraints and Opportunities.” Johannesburg, 18-20 October 2006.

Dustmann, C. (1994) “Speaking fluency, writing fluency, and earnings among migrants.” *Journal of Population Economics* 7(2): 133-156.

Dustmann, C. (1997) “The effects of education, parental background and ethnic concentration on language.” *The Quarterly Review of Economics and Statistics* 37: 245-262.

Dustmann, C. and van Soest, A. (2001) “Language fluency and earnings: Estimation with misclassified language indicators.” *The Review of Economics and Statistics* 83(4): 663-674.

Dustmann, C. and van Soest, A. (2002) “Language and the earnings of immigrants.” *Industrial and labour Relations Review* 55(3): 473-492.

Grin, F. (1994) “The economics of language: Match or mismatch.” *International Political Science Review* 15(1): 25-42.

- Grin, F. and Sfreddo, C. (1998) "Language-based earnings differentials on the Swiss labour market: Is Italian a liability?" *International Journal of Manpower* 19(7): 520-532.
- Heugh, K. (1999) "Languages, development and reconstructing development in South Africa." *International Journal of Educational Development* 19: 301 – 313.
- Heugh, K. (2000) "Multilingual voices – isolation and the crumbling of bridges." *Agenda* 46: 21-33.
- Heugh, K. (2007) "Language and Literacy Issues in South Africa". In: Rassool, N. (ed.) *Global Issues in Language, Education and Development*. Clevedon: Multilingual Matters Ltd., pp.187-217.
- Hofmeyr, J.F and Lucas, R. (2001) "The rise in union wage premiums in South Africa." *Labour* 15(4): 685-719.
- Kamwangamalu, N. M. (2002) "Language policy and mother-tongue education in South Africa: The case for a market-oriented approach". In Alatis, J.E., Hamilton, H.E. and A. Tan (eds) *Linguistics, language and the professions: Education, Journalism, Law, Medicine and Technology*. Washington: Georgetown University Press, pp.119-134.
- Kossoudji, S.A. (1988) "English language ability and the labor market opportunities of Hispanic and East Asian immigrant men." *Journal of Labor Economics* 6(2): 205-288.
- McManus, W., Gould, W. and Welch, F. (1983) "Earnings of Hispanic men: The role of English language proficiency." *Journal of Labor Economics* 1(2): 101-130.
- Mwabu, G. and Schultz, T.P. (2000) "Wage premiums for education and location of South African workers, by gender and race." *Economic Development and Cultural Change* 48(2): 307-334

- Pendakur, K. and Pendakur, R. (2002) "Language as both human capital and ethnicity." *International Migration Review* 36(1): 147-177.
- Probyn, M. (2009) "'Smuggling the vernacular into the classroom': conflicts and tensions in classroom codeswitching in township/rural schools in South Africa." *International Journal of Bilingual Education and Bilingualism* 12(2): 123-136.
- Probyn, M., Murray, S. Botha, L. Botya, P., Brooks, M. and Westphal, V. (2002) "Minding the gaps – an investigation into language policy and practice in four Eastern Cape districts." *Perspectives in Education* 20(1): 29-46.
- Rendon, S. (2007) "The Catalan premium: language and employment in Catalonia." *Journal of Population Economics* 20(3): 669-686.
- Shields, M.A. and Wheatley Price, S. (2002) "The English Language Fluency and Occupational Success of Ethnic Minority Immigrant Men Living in English Metropolitan Areas." *Journal of Population Economics* 15(1): 137-160.
- Stock, J.H.; Wright, J.H. and Yogo, M. (2002) "A survey of weak instruments and weak identification in generalised method of moments." *Journal of Business and Economic Statistics* 20: 518-529.