Globalization and the Returns to Speaking English in South Africa

by

James Levinsohn

University of Michigan
National Bureau of Economic Research

Current version: August 20, 2004

Address. Levinsohn: Ford School of Public Policy, University of Michigan, Ann Arbor, MI 48109; JamesL@umich.edu

Globalization and the Returns to Speaking English in South Africa

James Levinsohn
University of Michigan
National Bureau of Economic Research

1. Introduction.

The literature on globalization and wages is, by the standards of economics, huge. It is a literature that compensates for its volume by offering precious little in the way of convincing results. This is not (usually) the fault of the researchers. Rather, it is just very difficult to identify the role of international trade and/or investment on wages relative to the multitude of other factors that influence wages (and which frequently occur simultaneously with globalization.) This has led researchers to debate, for example, whether trade explains a growing wage gap between high wage and low wage earners or whether the real determinant of increasing wage disparity is coincident skill-biased technical change. Yet others (correctly) claim that even this dichotomy is a false one since international trade and investment and skill-biased technical change are themselves co-determined.

With this cacophony as background, this paper steps back and experiments with a very different approach to investigating the impact of globalization on wages. Noting the special circumstances around South Africa's emergence from the Apartheid era (and the relatively closed economy that accompanied the Apartheid era), this paper asks whether the return to speaking English (measured in a narrow way) increased as the South African economy embarked upon its integration with the rest of the industrialized world.

There is a certain logic to trying to measure the impact of globalization on wages in this manner. Following the advent of democracy in South Africa in 1994, there were several huge changes in the economy, many of which might be expected to change wages. One, but only one, of these changes was South Africa's re-integration with the global economy. Others included legislated changes in the labor market (with an emphasis on affirmative action) and the outbreak of the HIV/AIDS pandemic. Decomposing the changes in South African wages into those fractions due to increased disease, the dismantling of Apartheid and ensuing affirmative action, changes in technology during the 1990's, and increased integration into the global economy is a Herculean (or outright impossible) task. Measuring changes in the return to speaking English is a simple task and I argue below that

it is one that at least stands a chance of shedding light on the impact of globalization on wages in South Africa.

The underlying idea is that as South Africa re-integrated with the rest of the world, the return to speaking an international language of commerce might plausibly increase. In South Africa, English is that language. (The other widely spoken languages such as Zulu and Afrikaans are not used much in international commerce.) It is less obvious why some of the other changes concurrent with the fall of Apartheid should change the return to speaking English. It is, for example, unclear why AIDS should have much of an impact on the returns to speaking English (although it almost surely impacts wages.)[1] Nor is it clear why the sort of skill-biased technological change that occurred world-wide in the 1990s ought to impact returns to speaking English. Skill-biased technical change probably changes the returns to different levels of education but, conditional on education, it is hard to see why this sort of technical change would elevate the returns to speaking English. It is easier to suspect that affirmative action might impact the return to speaking English. This is a confounding influence that is explicitly discussed when presenting econometric specification and when interpreting results.

When substantial parts of the world did not openly trade with or invest in South Africa, there was still a return to speaking English. South Africa, after all, was not Albania. There remained some international trade, the mining industry produced traded goods, and there was some, albeit minimal, international investment in South Africa. Each of these might support a return to speaking English. Furthermore, speaking English was probably coincident with other factors that impacted wages given South Africa's history (see The Boer War.) For these reasons, this paper focuses whether that return to speaking English *changed*. Of course, if there is no return to speaking English in the first place, searching for changes in that return is not especially informative.

This approach to investigating the impact of globalization on wages is intended as a complement to the way economists usually address this question. My aims are pretty modest. This approach will not offer the definitive word on the impact of globalization on wages in post-Apartheid South Africa. Put another way, hard to imagine evidence on language as being dispositive. Nonetheless, when the cultural situation is appropriate, this approach might usefully add to the trade and wages debate (a.k.a. "cacophony."). Furthermore, this approach uses the sort of survey data that has for the most part been ignored in the trade and wages literature.

---

[1] One can of course concoct stories, some of them plausible, but few involve as direct a link between global integration and wages as that associated with the returns to speaking English

The paper proceeds in Section 2 by first describing some of the changes in openness in South Africa since the fall of Apartheid. Section 3 introduces the data that are used and provides some descriptive statistics. Section 4 estimates changes in the return to speaking English, while Section 5 concludes.

2. Background

In 1993, the first year of my data, South Africa was preparing for its first nationally representative election in decades. It was clear to all that a new government would be taking power in 1994. There was, though, considerable uncertainty regarding just what economic policies would be pursued by President Nelson Mandela. There were competing pressures to assure the international financial community of continued stability on the one hand, and to dramatically improve the lot of those who had for decades been excluded under the policies of the previous governments (and who were principally responsible for electing the new government) on the other hand.

South Africa quickly implemented a policy of macroeconomic stabilization to reassure the international financial community. Called GEAR for Growth, Employment, and Redistribution, the policy seemed to contribute to stabilization of key macro indicators such as inflation, real interest rates, and the budget deficit. It is less obvious that the policy enhanced growth, employment, and redistribution, but this of course depends on the counter-factual. Each component of the GEAR moniker might have been that much worse in the absence of the policy.

Encouraged by the sober fiscal policies of GEAR, companies from around the world that had hesitated before investing substantially in South Africa began to get off the sidelines. Foreign direct investment skyrocketed. Table 1 presents data for foreign direct investment in millions of Rand.

The data in Table 1 show that annual FDI inflows went from only 33 million Rand to over 1.3 billion as soon as the new government was ensconced and proceeded to increase to over 6 billion Rand by 2000. The huge inflow for the first half of 2001 is not typical and represents the one-off purchase of De Beers by the London-listed Anglo American Corporation. Even excluding that transaction, 2001 showed continued healthy increases in FDI inflows. According to the South African Reserve Bank, FDI was split pretty evenly between mining, manufacturing, and the financial sector.

South Africa also joined the WTO on January 1, 1995. Tariffs, never that high anyway, fell into the single digit range. The largest barrier to trade during the Apartheid era, though, was never tariffs. Rather, it was the willingness of the rest of the world to trade with South Africa. Under the new government, South Africa entered into regional free trading agreements with the European

Union and with the Southern African Development Community. Trade, as a percentage of GDP, increased substantially. Table 2 presents these figures. From 1991 to 1993, a period during which it became pretty clear that Apartheid was going to be replaced with a representative democracy, trade to GDP was pretty flat. It was with the new government in 1994 that trade as a fraction of GDP started to really increase. By 2000, the last year of my survey data, trade to GDP had risen almost 50 percent from .424 to .611. The ratio continued to rise and was .704 in 2002. By almost any standard, these are meteoric increases. Mirrored by the even greater increases in foreign direct investment, there is little doubt that the South African economy "globalized." South Africa clearly became more integrated with the global economy after 1993. I turn now to the question of whether the return to speaking English increased over the course of this period.

3. Data

This study uses data drawn from three South African household surveys-one from 1993 and two from 2000. The 1993 data are from the LSMS household survey conducted by The World Bank. This survey included about 44,000 individuals comprising just over 8800 households.[2] The version of the data often used by researchers contains about 300 variables.[3] Information on language and income are key variables for the study at hand. The data on language are not ideal due to the way that the survey instrument was worded. In particular, language is a household-level variable and the head of the household was asked to identify the "main language spoken at home." The fact that language is a household-level variable is not of particular concern, since the language spoken at home typically does not vary within the household. The fact that there is no information on whether a person could speak English instead of whether it is the main language spoken at home is a cause for concern and the results presented below must be considered in light of this. One would of course like to know whether one could speak English and how well, not whether it was spoken at home.

This is an example of one-sided measurement error. Some of those who are reported as not speaking English (as measured by the language spoken at home) in fact can speak English quite fluently. On the other hand, few or none of those who stated that English was their first language were in fact unable to speak English. This is because the answer to the language question was asked

---

[2] A cleaned and ready-to-use version of the data set, along with a primer to analyzing household survey data in STATA and the survey instruments are available at http://saproject.psc.isr.umich.edu/.

[3] The original data set includes over 2000 variables although many of these are essentially individual-level variables that are easily aggregated. Researchers who have used this data include Case and Deaton (19xx), Thomas (19xx), Duflo (19xx), and Lam (19xx) among others.

at the outset and determined the language in which the survey was administered. For example, if someone who spoke only Zulu stated that English was their language, that individual would have to then complete a multi-hour survey in English. It would not be hard to detect the mis-statement of language in this instance.

Using the available data probably lowers the estimated returns to speaking English. Assuming the survey under-counts those who speak English, then the true supply of English speakers is greater than the measured supply. Income data reflect the true supply so the estimated returns to speaking English are going to be lower. It may be helpful to consider the limiting case. Suppose that in reality almost everyone spoke English, just not at home. Then the estimated return to speaking English will be about zero.

The 1993 data on income are pretty good. The measure I use in this paper is an individual's total monthly income and is a constructed variable comprised mostly of wage income. It is common in developing countries to highlight the importance of accounting for self-production of food to properly compute income, but this is not an issue in South Africa. Own production is negligible.

For the 2000 data, I combine two surveys, the September 2000 Labour Force Survey (LFS) and the 2000 Income and Expenditure Survey (IES).[4]

Although the surveys are not explicitly linked, it turns out that the same households were included in both. The merged surveys result in a data set with about 101,000 individuals comprising about 26,000 households. The language question is in the 2000 LFS while income comes from the 2000 IES. The wording of the language question is the same as 1993. It again asks about the language spoken most often at home. The 2000 individual income data are used to compute total individual income in a manner most comparable with the 1993 definition. This amounts to subtracting various grants and pensions from individual income.

Table 3 presents frequency counts of language by race for each year of the sample. The sample is taken only among those reporting positive income and between the ages of 20 to 60. The lower bound is intended to exclude students and results are robust to a lower bound of 25 years instead of the somewhat arbitrary 20. The advantage of using 20 years old instead of 25 is that the sample size increases substantially. The upper bound is intended to exclude those receiving old-age pensions since those clearly do not depend on language spoken. Also, old age pensions are not going to be impacted by globalization as wage income might be. Some women begin to collect these pensions at

---

[4] There was also an Income and Expenditure Survey in 1995 and a linkable household survey (the October Household Survey). The 1995 IES and OHS are an attractive data source since questions are asked in the same way and one can be comfortable that income is measured consistently across the 1995 and 2000 surveys. Alas, the 1995 survey "forgot" to include the standard question on language.

age 60, hence the upper bound. There are three key messages from Table 3. First, very few Blacks[5] list English as their primary language. This is especially true in 1993 and it suggests that the return to speaking English within Blacks is going to be identified off of precious few observations. Second, Coloureds and Whites have a substantial numbers of English speakers. For each, Afrikaans is the majority language and for each there are substantial shifts in the fraction of the population group reporting English as their primary language. That fraction declines for Coloureds and increases for Whites. Third, English is essentially co-linear with Indian so that it will not be possible to separately identify the impact of English from the impact of being Indian on wages.

4. The Return to Speaking English

The question at hand is whether the return to speaking English (as imperfectly measured) increased as South Africa opened up to the international economy from 1993 to 2000. The return to speaking English is not directly observable and so needs to be inferred from econometric evidence. The approach adopted here is to estimate Mincer-like wage regressions and include as an explanatory variable whether the wage-earner listed English as his or her primary language. While simple in principle, several issues arise in practice.

First, it is necessary even in the cross-section to include as explanatory variables key determinants of wages.[6] Omission of an explanatory variable that itself might be correlated with speaking English will bias the estimate on the return to speaking English. Second, the many changes in South Africa from 1993 to 2000 probably impacted many of the determinants of wages. It is widely believed, for example, that the return to education and the wage differentials apparently due to race changed over this period. Holding them constant and only allowing the return to English to change will yield biased estimates of the true change in the return to speaking English. (On the other hand, such an approach pretty much guarantees finding a pretty big change in the return to speaking English.) Third, the fact that about 40 percent of English speakers are Indian and there is virtually no language variation within this population group poses a challenge. The most flexible approach to estimating the returns to speaking English examines the change in that return within population group yet this approach is going to be non-informative for Indians.

---

[5] I use the term "Blacks" since this seems to be preferred by most South Africans to the term "Africans" that is used in the survey instrument. For data purposes, the two terms are interchangeable.

[6] A difference in differences approach is ill-advised because of concurrent changes in many other variables impacting wages. That is, while one could measure the difference in wages between those who list English as their first language and those who do not, and one could then examine the difference over time in this difference, the result would be hard to interpret. This is because many other variables changed over this period and some of those changes are not orthogonal to an observed return to speaking English.

The simplest specification regresses log individual income (y_i) on indicator variables for each value of j years of education (ED), experience (EX), experience squared, an indicator variable for whether the worker is male (M), indicator variables for population group (CO–Coloured, IN– Indian, and WH– White, with Blacks as the excluded group), and an indicator for whether English is the language spoken at home (ENG). Experience is defined as age minus 20. Hence,

$$lny_i = \beta_0 + \sum_{j=2}^{j=13} \beta_{1,j} ED_j + \beta_2 EX + \beta_3 EX^2 + \beta_4 M + \beta_5 CO + \beta_6 IN + \beta_7 WH + \beta_8 ENG + \epsilon_i \quad (1)$$

Equation (1) is estimated separately for each year of the sample using Ordinary Least Squares (OLS) with the appropriate sample weights. Estimating the regression separately for each year is necessary to capture the changes in returns to education between 1993 and 2000 as well as changes in the return to being male and/or of a particular population group. Use of indicator variables for each level of education permits returns to vary non-linearly with years. The coefficient on English, β_8 is interpreted as the percentage wage differential attributable to speaking English conditional on the other included regressors. The results from this specification applied to the 1993 and 2000 data are presented in Table 4.

The results from 1993 are discussed first to fix ideas. The first 12 rows show the usual returns to education. For example, someone with 12 years of education, all else equal, earns about 168 percent more than those with one year or less of education conditional on the other co-variates. The wage premium for being a member of a race other than Black ranges from 32 percent for Coloured to 98 percent for White. Males earn 46 percent more than similar females. The coefficient of interest for this study, though, is that on "English." Conditional on education, experience, gender, and race, people who list English as their primary language earn about 18 percent more than those who list another language. This differential is quite precisely estimated.

Equation (1) is estimated using the 2000 data and the results are in column two of Table 4. While there are several interesting comparisons between 1993 and 2000 to be made (the changing pattern of the return to education for instance), the focus here is on the impact of speaking English. The "English premium" jumps from .183 in 1993 to .252 in 2000. The 2000 premium is precisely estimated and the change between the two years is significantly different than zero. Allowing the entire pattern of returns to years of schooling to vary from 1993 to 2000, and allowing for differing returns to race, gender, and experience, it is still the case that the return to speaking English increased substantially. This change in the English premium as well as its level are of

7

an economically large magnitude. By 2000, English speakers were earning about 25 percent more after conditioning on other observables and the premium had increased by 7 percentage points since 1993.

The specification reported in Table 4 imposes that the returns to education, experience, and gender are identical across racial groups. A convincing body of research suggests this is too strong an assumption. I proceed by looking for the English premium within each of the racial groups. Doing so allows the returns on all the other observables to vary by racial group. This flexibility is clearly a good thing for it lets the data speak more "freely." The flexibility, though, will carry a price. Thirty to forty percent of the sample that listed English as the language spoken at home are Indian, and virtually all Indians list English as the primary language. There is, then, no within group language variation for Indians. Hence, it is not feasible to estimate a return to speaking English for Indians since that return is not identifiably different than the return to simply being Indian.[7]

Table 5 reports results from the within-group regressions for Blacks, Coloureds, and Whites. In the interest of parsimony, only the coefficient on speaking English is reported.[8] This approach is pretty flexible. It allows the returns on all observables to vary both over time and across racial groups.

As is usually the case with a more flexible specification, the messages are more mixed than those reported in Table 4. The first row of Table 5 reports the English premium from Table 4 for comparison's sake. The next three rows report the English premium for the other racial groups (except Indian for reasons discussed above.) For Blacks, the English premium stayed constant from 1993 to 2000. It was huge (about 60 percent) but did not increase over time, although the precision of the estimate did increase. One should recall, though, that this premium is being identified off of very few individuals– 6 out of 2468 in 1993 and 118 out of 20,222 in 2000. For Coloureds, the return to speaking English fell about 11 percentage points. The decline, while not large, is statistically significantly different than zero. The largest change from 1993 to 2000 in the English premium impacted Whites. The return went from being basically non-existent in 1993 to a precisely estimated 14.5 percent. Another way to interpret this result is that the "penalty" to speaking Afrikaans among Whites skyrocketed.

---

[7] It is possible to estimate a return to speaking English among Indians, but the effect is identified off of 3 individuals who listed "Other" in 1993 and about 6 Afrikaans-speaking Indians in 2000. The English premium, when separate regressions are run for Indians, is never significantly different than zero.

[8] Were it the case that the returns on the other observables (not reported in this table) did not vary significantly across racial groups, it would be efficient to pool groups. Alas, coefficients vary across groups and one can readily reject the hypothesis that the returns to observables other than the English premium are the same across groups.

The general pattern reported in Table 5 is robust to many alternative specifications. For example, the inclusion of indicator variables for the province in which a household lives, using a single variable for years of education instead of the more flexible set of indicator variables, the exclusion of the variable for male, using age 25 as a lower age bound, using 65 as an upper age bound, and interacting the return to English with education all yield basically the same message when it comes to the English premium. Namely, that premium became much larger for Whites, fell slightly for Coloureds and for Blacks.[9] These results, though, exclude a large number of those with positive income for whom English is the first language– Indians.

5. Concluding Remarks

Did globalization *really* cause the return to English to increase in South Africa? The evidence in this paper is, in some cases, corroborating, but hardly conclusive. The strongest and most robust result is the the return to speaking English increased for Whites over the period during which South Africa re-integrated with the world economy. This result strong because it results from the flexible within-group estimates, and it is robust because it arises in all the investigated specifications. When Indians are included and a (necessarily) less flexible estimation strategy is adopted, I again find that the return to English increased and that the increase is precisely estimated. These are the results in Table 4. If one thinks of these results as indicating an "average" effect of speaking English, that effect is positive. There is less evidence, though, that the return to speaking English increased among Blacks and Coloureds.

One explanation for the lack of an increase in the return to speaking English among Blacks is the following. In 1993, there were few Blacks that spoke English and they earned a premium for their language skills. With the advent of affirmative action, the premium for speaking English fell as more Blacks were promoted into higher paying jobs. In this case, it was no longer just the few English speaking Blacks earning the relatively higher wages. This scenario illustrates one of the difficulties of disentangling the impact of globalization (which might actually increase the return to speaking English) with the impact of affirmative action for Blacks (which was concurrent with globalization and which might actually decrease the "extra" return to speaking English.)

There was no affirmative action for Whites and among this group, the return to speaking English clearly rose. Put another way the penalty for speaking Afrikaans rose for Whites. This is consistent with capturing an impact of globalization. Afrikaans, unlike English, is much less useful

---

[9] For some specification, the premium rises slightly for Blacks.

in international commerce. Those Whites whose first language was English benefited conditional on education, gender, and experience. This is, as noted above, corroborating but not conclusive evidence.

The finding that the return to speaking English did not increase for Coloureds muddles the waters. Coloureds did not benefit from affirmative action as did Blacks under the new government. Still, the return to speaking English did not rise and in fact fell. If globalization is what moves the return to speaking English, one should have found an increase to speaking English among Coloureds and this was not the case.

The evidence in the end is mixed. On the whole, the return to speaking English increased but within racial groups, the pattern is not consistent.[10] The approach adopted in this paper is perhaps a novel way to revisit the wages and globalization issue. It is an approach that is especially well-suited to developing countries, many of which have a rich variety of languages spoken, as they integrate with the global economy. In other contexts (India, for example), or with better data (industry of employment data, for example), the approach adopted here may prove more conclusive. Or not. Even if language is an accurate way to isolate an impact of globalization on wages, it may simply be that globalization has differing impacts on differing segments of a population. This appears to be the case in South Africa.

---

[10] It should be noted that precious few of the English speakers are among the very poor. In 1993, virtually none are while in 2000 only a handful are. Hence, this approach does not speak to the role of globalization on the incomes of the very poor.

References

[TO BE COMPLETED.]

	TABLE 1 Foreign Direct Investment in South Africa ('000,000 Rand)
year	FDI
1993	33
1994	1348
1995	4502
1996	3515
1997	17587
1998	3104
1999	9184
2000	6083
2001*	53000

Notes: Source is the OECD Global Forum on International Investment. All figures are nominal millions of Rand. * indicates data are only for the first half of 2001.

TABLE 2 Trade / GDP in South Africa	
year	
1991	.440
1992	.423
1993	.424
1994	.452
1995	.48
1996	.514
1997	.520
1998	.550
1999	.541
2000	.611
2001	.656
2002	.704

Notes: Source is www.resbank.co.za. Data are for imports and exports of goods and services and annual GDP.

TABLE 3 Primary Language Among Wage Earners 20 to 60 Years Old					
1993					
Language	Black	Coloured	Indian	White	Total
English	6	160	232	202	600
Afrikaans	11	411	0	522	944
Xhosa	421	0	0	0	421
Zulu	624	0	0	0	624
Tswana	493	0	0	0	493
N. Sotho	285	0	0	0	285
S. Sotho	91	0	0	0	91
Venda	79	0	0	0	79
Tsonga	209	0	0	0	209
Swazi	173	0	0	0	173
Ndebele	69	0	0	0	69
Other	7	0	3	6	16
Total	2,468	571	235	730	4,004
2000					
English	118	534	626	783	2,075
Afrikaans	230	3,032	6	1,540	4,838
Ndebele	435	0	0	0	435
Xhosa	3,789	17	0	0	3,806
Zulu	5,487	3	0	0	5,490
N. Sotho	2,238	1	0	0	2,239
S. Sotho	2,681	6	0	1	2,688
Tswana	2,601	17	0	0	2,618
Swazi	871	4	0	0	876
Venda	577	0	0	0	577
Tsonga	1,109	0	0	0	1,109
Other	72	2	22	19	120
Missing	14	0	0	0	14
Total	20,222	3,616	654	2,343	26,885

Note: 2000 row totals do not sum properly due to the exclusion of non-responses to the race question.

	1993	2000
TABLE 4 The Returns to Speaking English 20 to 60 Years Old		
ed2	-.093 (.071)	.092 (.035)
ed3	.219 (.067)	.233 (.038)
ed4	.299 (.065)	.224 (.036)
ed5	.449 (.060)	.295 (.034)
ed6	.472 (.052)	.380 (.031)
ed7	.733 (.051)	.549 (.030)
ed8	.622 (.060)	.646 (.032)
ed9	.948 (.050)	.820 (.030)
ed10	.993 (.063)	.908 (.031)
ed11	1.274 (.047)	1.240 (.027)
ed12	1.688 (.054)	1.796 (.031)
ed13	1.788 (.075)	2.126 (.036)
EX	.062 (.004)	.099 (.002)
EX2	-.001 (.000)	-.001 (.000)
Coloured	.326 (.037)	.360 (.020)
Indian	.394 (.071)	.421 (.041)
White	.984 (.037)	.921 (.021)
English	.183 (.043)	.252 (.024)
Male	.463 (.024)	.501 (.012)
Constant	5.041	7.183

	TABLE 5 The Returns to Speaking English Within Group Results	
	1993	2000
All	0.183 (0.043)	0.252 (.024)
Black	0.592 (0.290)	0.380 (.084)
Coloured	0.521 (0.0765)	0.410 (.042)
White	-0.017 (.050)	0.145 (.036)