

The Returns to English-Language Skills in India

Mehtabul Azam

Aimee Chin

Nishith Prakash

World Bank and IZA*

University of Houston and NBER†

Dartmouth College and IZA‡

September 2009 Draft: PRELIMINARY AND INCOMPLETE §

Abstract

India's colonial legacy and linguistic diversity give English an important role in its economy, and this role has expanded due to globalization in recent decades. It is widely believed that there are sizable economic returns to English-language skills in India, but the extent of these returns is unknown due to lack of data containing measures of both earnings and English ability. In this paper, we use a newly available data set—the India Human Development Survey, 2005—to quantify the effects of English-speaking ability. We find that being fluent in English (compared to not speaking any English) increases hourly wages by 32%, which is as much as the return to completing secondary school and half as much as the return to completing a Bachelor's degree. Being able to speak a little English significantly increases hourly wages 12%. We find strong evidence of language-skill complementarity, with more experienced, educated and able workers receiving a higher return to English-language skills.

JEL Classification: J31, J24, O15

Keywords: English Language, Human Capital, India

*Consultant, World Bank, Washington, DC 20433 USA (e-mail: mazam1@worldbank.org).

†Department of Economics, University of Houston, Houston, TX 77204-5019 USA (e-mail: achin@uh.edu).

‡Department of Economics, Dartmouth College, Hanover, NH 03755-3514 USA (e-mail: nishith.prakash@dartmouth.edu).

§We thank Eric Edmonds and Chinhui Juhn for helpful comments and discussion. The findings and opinions expressed here are entirely those of the authors and do not necessarily represent those of the World Bank.

1 Introduction

India's colonial legacy and linguistic diversity give English an important role in its economy, and this role has expanded due to globalization in recent decades. It is widely believed that there are sizable economic returns to English-language skills in India, but the extent of these returns is unknown due to lack of data containing measures of both earnings and English ability.¹ We take advantage of a newly available nationally representative individual-level data set, the India Human Development Survey (IHDS), 2005, to provide the first estimates of the returns English-language skills in India. A second contribution of this paper is to provide new descriptive information about the prevalence of English ability in India. Based on the 1991 Census, 11% of the Indian population reported English ability. It would be useful not only to have more recent figures, but also to examine English ability along various dimensions such as education and sex.

A major challenge to estimating the returns to English is the likely endogeneity of language skills in the earnings equation. Omitted variables bias, such as omitted ability or local labor market conditions, is one concern. We exploiting the size and richness of IHDS data to address these concerns; for example, the IHDS has data on individual performance on the secondary school leaving certificate board examination, which provides a credible proxy for ability.

Approaching this question using India as a setting is very informative for a number of reasons. First, India is one of the world's most linguistically diverse country, in addition to being one of the countries where English was established firmly as the medium of instruction and administration by the British Raj as early as 1765. Approximately more than 180 distinct languages are spoken in India. Second, in terms numbers of English speakers, about 35 million people (in 1994) use English. Hence the results drawn from this study have

¹Munshi and Rosenzweig (2006) and Chakraborty and Kapur (2008) estimate the returns to English-medium education, which we argue in subsection 2.2 is not the same as the returns to English-language skills.

broader implications. Also we are not aware of any study that estimates the returns to English-language skills in India that is nationally representative.

Our main findings are as follows. First, there are large, statistically significant returns to English-language skills in India. We find that being fluent in English (compared to not speaking any English) increases hourly wages by 32%, which is as much as the return to completing secondary school and half as much as the return to completing a Bachelor's degree. Being able to speak a little English significantly increases hourly wages 12%. We find strong evidence of language-skill complementarity, with more experienced, educated and able workers receiving significantly higher returns to English-language skills.

The remainder of this paper is organized as follows. Section 2 provides background on English in India and discusses the related literature. Section 3 presents the empirical framework. Section 4 describes the data. Section 5 presents the results on returns to English-language skills in India. Section 6 provides a discussion and policy implications.

2 Background and Related Literature

2.1 English in India

India is a linguistically diverse country—it has thousands of languages, of which 122 have over 10,000 native speakers according to the 2001 Census. English is only 44th on the list of languages in India with the most native speakers, belying its important role in India since the arrival of the British East India Company in the 1600s. India was formally ruled by the British Empire from 1757-1947 (by the British East India Company from 1757-1857, and by the British Crown from 1858-1947). During this time, English became the language of power and prestige. It was associated with the ruling British, the law was in English, and government administration, at least at the higher levels, was conducted in English.

Additionally, it became the medium of instruction in public schools.²

After India gained independence from the British in 1947, debate ensued over the role of the colonial language in the country. There were calls to replace English with a native Indian language as the official language of India to reinforce national identity. A natural candidate was Hindi, which is by far the most dominant mother tongue in India (in 2001, 40% of the population named Hindi as their mother tongue; the next language with most native speakers, Telugu, claimed 8%). However, it was politically infeasible to make Hindi the sole official language of India as it was thought to be disadvantageous to states where Hindi was not prevalent—Hindi is spoken by most in the north, by few in the south. Thus, the Constitution of India names both Hindi and English as the official languages of India. Individual states legislate their own official languages, but communication among states and in the federal government would take place in Hindi or English.

From an individual's perspective, there are several economic incentives to learn English. On the one hand, English has value as a lingua franca. A knowledge of a common language facilitates communication. A common language is especially useful in linguistically diverse places, where the chances of meeting someone with the same native language is relatively low. In India, there is considerable variation in languages spoken even within narrowly defined regions, such as the district.³ A common knowledge is also useful for international trade. While English is not the only possible lingua franca, it is a natural one given India's colonial past and given the influence of the United States in the world economy. On the other hand, the use of English is firmly entrenched in government and schools due to the colonial past. To be a government official or teacher (other than at low levels), one needs to be proficient in English. These occupations are considered attractive in India because they are white-collar jobs providing secure employment and good benefits. In contrast, most jobs in the India are

²Under British rule, India established a system of public education; before, there were few schools and only the elite received schooling. It was decided after much debate that English would be the medium of instruction in this new system of public schools.

³See Shastri (2008) for more on within-district variation in languages spoken. India is comprised of 35 states/union territories which are subdivided into 628 districts.

on household farms or in casual labor, which tend to provide uncertain means of livelihood and involve strenuous physical labor.

Though only 0.2% of the Indian population reported English as their mother tongue in the 2001 Census, considerably more know it as a second or third language. According to the 1991 Census, 11% of the Indian population reports English as a second or third language. It is widely believed that English knowledge has grown since 1991, but there has been no data to substantiate these claims until now, with the release of the India Human Development Survey (IHDS), 2005 (we describe these data in Section 4). Table 1 reports the mean English ability among individuals aged 18–65 in the IHDS along various dimensions. One in five Indians report having the ability to speak English. Out of this 4% can converse fluently in English while 16% can converse a little in English. English ability is higher among men—approximately 26% of men report having the ability to speak English compared to 14% of women—and this is likely due to the differences in educational attainment, which we discuss below. English ability is higher among younger people—25% of people aged 18-35 speak English compared to 13% for people aged 51-65. These differences may be due to differences in educational attainment, greater incentives to learn English due to globalization in recent decades, or depreciation of English skills with time since leaving school.

The ability to speak English increases dramatically with educational attainment in India. Almost 89% of individuals who have at least a Bachelor’s degrees can speak English as compared to 56% for those who have completed 10-12 years of schooling, 11% for those who have completed 5-9 years, and virtually nil for those who have less schooling. The positive relationship between English ability and educational attainment is not surprising since English is not the native language of 99.8% of the Indian population, and thus the main exposure to English for children would be in schools. In India, many public schools follow the “Three Language Formula” recommended by the central government, which generally leads to teaching in English by middle school.⁴ According to the 1986 All-India Education Survey,

⁴This calls for the teaching in the mother tongue or regional language during primary school. After

which is a census of schools, 1.3% of schools with grades 1-5 used English as the medium of instruction, and 15% reported teaching English as a first or second language. In schools with grades 6-8, these figures rise to 3.6% and 63%, respectively. In secondary schools (covering grades 9 and 10), 8.2% use English as the medium of instruction, and 65% teach English as a first or second language. In higher secondary schools, colleges and universities, English is often used though it should be pointed out that it is possible to graduate from secondary school and college without being proficient in English—except in the science and engineering fields, many courses are offered in Hindi or the state language or Hindi, and exams may be written in English, Hindi or the state language.

Continuing with Table 1, among social groups, more individuals belonging to higher castes speak English as compared to individuals from lower castes or the scheduled tribes. This is likely related to the lower educational attainment, and in the case of the scheduled tribes, of their geographic isolation.

The bottom of Table 1 presents mean English ability by geography. There is a large difference in English ability by urban status: 34% of individuals living in urban areas report to have ability to speak English as compared to only 14% living in rural areas. There is considerable variation in English ability across the regions as well. This is further illustrated in Figure 1, which displays mean English ability by state.

Having documented the prevalence of English proficiency in India using the IHDS, we proceed to estimate the returns to English proficiency in India. Before we do this, we discuss the related literature and our empirical framework.

2.2 Previous Literature

We are aware of two previous studies on the relationship between English-language skills and earnings in India: Munshi and Rosenzweig (2006) and Chakraborty and Kapur (2008),

primary school, introduce a second language—this might be Hindi (in states where Hindi is not the dominant language) or English or some other modern Indian language. After middle school, introduce a third language.

where the latter is an unpublished manuscript. Both estimate the returns to attending a school with English (as opposed to the native language) as the medium of instruction. Munshi and Rosenzweig collected their own data on Maharashtrians living in Dadar, which is a place in Mumbai, Maharashtra, India. Using data on parents' income histories and the language of instruction in their secondary school (Marathi or English), they estimate significant positive returns to an English-medium education.⁵ Attending an English-medium school increased both women's and men's income by about 25% in 2000. Chakraborty and Kapur use National Sample Survey data to estimate the impact of a 1983 policy in West Bengal which eliminated English as the medium of instruction in primary schools. They find that switching from English to Bengali medium of instruction significantly reduced wages. Simple comparisons of cohorts attending primary school before and after the policy change suggest that English-medium schooling raised wages about 15% in the 2000s.⁶

Our study differs from the two aforementioned studies in two key respects. First, the “returns to English” that we are estimating is the returns to English-language skills as opposed to the returns to English-medium education. In general, we might think that being taught in English would increase one's English-language skills relative to being taught in some other language, so the latter estimates just need to be scaled up by some factor to obtain the former.⁷ Angrist and Lavy (1997), for example, find that French-language skills significantly deteriorated in Morocco as a result of a policy that changed the language of instruction in post-primary grades from French to Arabic. However, Angrist, Chin and Godoy (2008) find that in Puerto Rico, switching the medium of instruction from English to Spanish in Puerto Rico had no impact on the English-speaking proficiency of Puerto Ricans; thus, it is not a

⁵These returns are described in greater detail in Munshi and Rosenzweig (2003).

⁶Estimates controlling for secular cohort trends suggest somewhat larger effects.

⁷That is, we might think of the returns to English-medium schooling as a reduced-form relationship between English-language skills and earnings. In order to recover the returns to English-language skills, one needs to know the “first-stage” effect of English-medium schooling on English-language skills. In practice, there might be complications since English-medium schooling might impact earnings through mechanisms other than English-language skills. For example, Roy (2004) finds that the West Bengal policy increased educational attainment.

foregone conclusion that instruction in a foreign language will lead to greater proficiency in that foreign language. In fact, the premise of He, Linden and MacLeod (2008) is that Indian primary schools are ineffective at teaching English.⁸ A second difference is that our study uses a nationally representative data set, which enables us to explore potential heterogeneity in returns to English-language skills by region and along other dimensions. The Munshi and Rosenzweig study is focused on one community in Mumbai, and Chakraborty and Kapur (2008) are focused on West Bengal.

There is a large literature on the effects of language skills on wages using data from other countries, however. Most of these studies estimate the returns to the host-country language for immigrants to that host country, such as the returns to English for U.S. immigrants. Bleakley and Chin (2004) provide a brief overview of these studies. Fewer studies estimate the return to a language that is not the country's dominant language. Two studies that estimate the effect of a colonial language are Angrist and Lavy (1997), who estimate the return to French-language skills in Morocco, and Levinsohn (2007), who estimates the returns to speaking English in South Africa. Two that estimate the effect of foreign languages that do not have a colonial past in the country are Saiz and Zoido (2005) and Lang and Siniver (2006). Saiz and Zoido estimate the returns to Spanish, French and other foreign languages among U.S. college graduates. Lang and Siniver estimate the returns to English proficiency in Israel, a country where English is neither a dominant nor official language. The latter two studies suggest that proficiency in an international language such as English is rewarded more in the labor market than proficiency in some other foreign language.

Our study is also related to a couple of recent unpublished manuscripts on the connection between language and economic development in India. First, Shastry (2008) finds that Indian districts whose population's mother tongue is more linguistically dissimilar to

⁸They perform a randomized evaluation of a new methodology for teaching English in primary schools. At the outset of the experiment, they found that only 10% of second and third graders could identify the picture of the correct object when given the object's English name even though these words were part of the official English curriculum.

Hindi to attract more information technology jobs, which she attributes to the lower cost of learning English.⁹ Then she finds that a greater IT presence is associated with a greater increase in school enrollment and a smaller increase in the wage premium for educated workers in districts where the mother tongue is more linguistically dissimilar to Hindi. However she does not have individual-level data on English-language skills—her language variables are at the district level—and does not estimate returns to English proficiency per se. Second, Clingingsmith (2008) finds that districts that had greater increases in manufacturing employment experienced greater increases in the proportion of minority-language speakers becoming bilingual (where the second language is a regional or national language).

This study makes several contributions. It is the first to estimate the returns to English-language skills in India. A major hurdle in the past has been the availability of a data set measuring both language skills and earnings of individuals. The recent availability of data from the India Human Development Survey (IHDS), 2005 clears this hurdle for us. Additionally, it adds to the handful of studies that estimates the effect of proficiency in a foreign language that has colonial roots or serves as an international language. Finally, it adds to the few studies on the topic using data from a developing country.

3 Empirical Framework

English-language skills are a form of human capital. Individuals, or parents acting on their behalf, weigh the marginal costs and marginal benefits of investing in English-language skills. There could be both monetary and non-monetary costs associated with acquiring English-language skills. Non-monetary costs include the effort to learn English, which is not the native language of 99.8% of the Indian population. They might also include weakened ties to one's tradition social network because in order to learn English well, one might have to attend different schools or study the native language less relative to other members of the

⁹The reasoning is that people whose native language is not Hindi or English will learn Hindi if their mother tongue is very similar to Hindi and English otherwise.

network (Munshi and Rosenzweig 2006). Extra monetary costs are incurred to the extent that parents enroll their child in private schools, hire tutors, or invest in more years of schooling than they otherwise would in order to help the child learn English.¹⁰

Among the benefits of having English-language skills is getting a better job. Though it is widely acknowledged that English-language skills are valuable in India, due to lack of data, the returns to English-language skills have not been quantified.¹¹ This skill price is determined by the supply of and demand for English-proficient workers. The recent availability of data from the India Human Development Survey (IHDS), 2005, which contains measures of English-speaking ability and earnings, enables us to estimate it.

The relationship between English-language skills and earnings might be approximated by the following equation:

$$y_i = \alpha + \beta English_i + \pi X_i + e_i \quad (1)$$

where y_i is the log earnings of individual i , $English_i$ is a measure of English-language skills, and X_i is a set of demographic controls (e.g., age, sex, social group). The coefficient of primary interest is β , which gives the returns to English-language skills.

Omitted variables bias will likely be a concern for several reasons. One omitted variable that ex ante we thought would be important is years of schooling. Early grades are more likely to take place in the native language, with the regional or national language used as the medium of language in later grades. This generates a positive relationship between English proficiency and years of schooling, leading the ordinary least squares estimate of β in Equation 1 to be upward biased; some of the estimated effect of English proficiency is in fact due to schooling. To address this issue, we will control for years of schooling.

¹⁰Roy (2004) finds that as a result of the change in medium of instruction in West Bengal public primary schools from English to Bengali in 1983, parents spent more on private tutors (presumably to provide English lessons). There was no estimated impact on private school attendance but Roy suggests that perhaps there were supply constraints in the short run.

¹¹As discussed in the previous section, Munshi and Rosenzweig (2006) and Chakraborty and Kapur (2008) estimate the returns to English-medium schooling which is not the same as the returns to English-language skills.

A second important omitted variable in Equation 1 is geographic characteristics. Places where English is more prevalent are different from places where English is less prevalent. For example, English is more prevalent in major cities and outside the Hindi belt (states in central and northern India where Hindi is the mother tongue), but these are also places where wages are on average higher. There could be a causal relationship between wages and English-language skills, or a correlation could exist through some third factor. A causal story going from language skills to industrial development is that the availability of a more skilled (e.g., English proficient) labor force in a particular region attract firms to locate there. Shastri (2008) finds that districts with a lower cost of acquiring English (and therefore, where English was more prevalent) had greater growth in information technology jobs. A causal story in the reverse direction is that higher returns for a particular skill (e.g., English proficiency) motivates individuals to invest more to develop that skill. Clingingsmith (2008) finds that Indian districts with more factory employment experience greater growth in bilingualism. Third factors which might be behind the observed correlation include the place's institutional environment, economic conditions, climate and natural resources. The foregoing considerations suggest that it is essential to control for local conditions. We therefore control for urban residence and state of residence. In some regression models, we control for geographic conditions more finely by including district fixed effects, i.e., we are comparing individuals who speak English and those who so not *within* the same district.

A final important source of omitted variable bias in Equation 1 is the standard “ability” bias. Higher-ability individuals (e.g., those with greater aptitude or more advantaged family background) are more likely to have better English-language skills as well as better jobs, leading to an upward biased estimate of the return to English since ability is omitted. Controlling for years of schooling mitigates the ability bias somewhat because those with greater ability tend to complete more schooling, but this is imperfect. We take advantage of the richness of the IHDS data to address the ability bias in additional ways. One way is to control for father's education; controlling for parental education has been a traditional approach

in the returns to education literature to address ability bias (see Card(1999) for a review). Another way is to control for the individual’s performance on the secondary school leaving certificate (SSLC) examination. In India, students must pass a standardized exam developed by the board of education under whose jurisdiction their school falls in order to receive a SSLC. This exam is typically taken at the end of 10th grade, and the passing categories, from highest to lowest level of distinction, are Division I, II and III. Performance on the SSLC exam—even if measured in a only a few categories—seems like a reasonable proxy for ability, and is akin to controlling for aptitude test scores to address the ability bias when estimating the returns to schooling. It is unusual, and fortunate for us, that the IHDS collected this SSLC exam performance data. However, it should be recognized that this information is available only for individuals who have attended 10th grade, something that only a fifth of individuals aged 18-25 in the data set has done. Thus we regard SSLC exam performance as a proxy of ability among those with more schooling. As a proxy among the less educated, we use the response to the question of whether the individual has failed or repeated a grade.

Given the forgoing discussion on likely sources of omitted variables bias, we amend Equation 1 as follows:

$$y_i = \alpha_r + \beta English_i + \delta Schooling_i + \gamma AbilityProxies_i + \lambda Urban_i + \pi X_i + e_i \quad (2)$$

where $Urban_i$ indicates whether individual i lives in an urban area, α_r is either state or district fixed effects, $Schooling_i$ is years of schooling completed and $Ability_i$ is a proxy for ability (i.e., father’s education, SSLC exam performance, and failing or repeating a grade).

Though we believe Equation 2 addresses the main sources of bias, nonetheless there could be concerns about remaining omitted variables and measurement error. Measurement error is potentially serious because respondents self-report their English ability and English ability is measured in a few categories. Previous work addressing the endogeneity of language-skills measures in the earnings equation suggests that the downward bias due to classical

measurement error may be much more severe than the upward bias due to ability bias (Bleakley and Chin (2004), Dustman and van Soest (2002)), and consequently instrumental-variables estimates of the returns to English ability tend to be larger than ordinary-least-squares estimates (which would in general be affected by both measurement error and ability bias). In future work, we will search for an instrumental variables strategy for English-language skills; given the aforementioned studies, we might interpret the estimates here as lower bounds on the return to English.

4 Data

We use data from the 2005 India Human Development Survey, a nationally representative household data set collected by the National Council of Applied Economic Research in New Delhi and the University of Maryland (Desai, Reeve and NCAER 2009). IHDS covers 41,554 households located throughout India.¹² This data set became available for public use only recently, and has not yet been exploited by economists. The survey contained many questions that are not asked in the larger and more commonly used Indian household survey, the National Sample Survey. Most relevant for us is that information about each household member's ability to converse in English is collected. We are not aware of any other large-scale individual-level data set in India that contains a measure of English-language skills for everyone.¹³

Since the outcome of interest is earnings, we restrict our sample to individuals aged 18 to 65. Our main analysis will use individuals who report working for a wage or salary last year. In our sample of 125,170 people aged 18-65, 38% participated in wage employment. This raises concerns sample selection bias. This problem could be serious in the context of

¹²The survey covered all the states and union territories of India except Andaman and Nicobar and Lakshadweep, two union territories which together account for less than .05% of India's population.

¹³The census does collect information on language knowledge of the population, however these data are reported only in aggregate form. Additionally, the census does not collect data on wages, income or consumption.

India, where over 70% of the population is rural, and family farms and non-farm businesses continue to absorb much of the labor force. To address this, we will use two alternate measures of earnings that are observed regardless of an individual's employment status: household income and household consumption. We perform this latter analysis using male heads of households to avoid counting the same household multiple times. Table 2 presents the descriptive statistics for the wage sample and male household head sample.

5 The Returns to English-Language Skills

We present the results of estimating Equation 2 in Table 3, with each column adding additional sets of controls. Column 1 presents the raw difference in log hourly wage by English proficiency. Compared to log wages of 1.9 for individuals who have no English ability, those who speak a little English earn 82 log points more, and those who speak English fluently 149 log points more. These are very large differences but potentially overstate the returns to English due to our observations in Table 1 that English ability is correlated with numerous variables that are known to be correlated with earnings, such as age and education.

In Column 2 of Table 4, we control for sex, age and social group. In Column 3, we add a full set of dummies for educational attainment. Controlling for education cuts the estimated coefficients for English ability by more than half. Adding state fixed effects and a dummy for urban residence, which helps control for local institutions and labor market conditions, also leads to sizable reductions in the estimated coefficients for English ability (Column 4). In Column 5, we add dummies for SSLC exam performance and grade repetition/failure to control for ability bias (on top of what controlling for educational attainment accomplishes). Although performing in the top category (Class I) on the SSLC exam significantly increases wages, and having ever failed or repeated a grade significantly decreases wages, we find a relatively modest decline in the estimated coefficients. Finally in Column 6, we add district fixed effects, which is by far more detailed a geographic control than the state (India has

35 states/union territories which are subdivided into 628 districts) and this does not make much difference. The estimates in Column 6 suggest that, compared to having no English ability, being fluent in English increases hourly wages by 32%, and being able to speak a little English increases wages by 12%. These are economically meaningful effects. For example, the return to being fluent is as large as the return to completing secondary school, and half as large as the return to completing a Bachelor's degree. However, these returns are smaller than what has been estimated for the returns to English for U.S. immigrants.

An issue with our analysis thus far is that it is restricted to people engaged in wage employment, which may be a non-random sample of the population. We address this issue by looking at household income and consumption as outcomes; we have measures of income and consumption for all households regardless of members' participation in wage employment. We present this analysis in Table 4, which is restricted to male-headed households. In Column 1, we repeat estimate the same specification as in Table 3, Column 6 to demonstrate that very similar effects on wages are found when we use the sample of male household heads. In Column 3, we use household income as the outcome. We find that in households where the head speaks English fluently, annual income is 34% higher. In households where the head speaks a little English, annual income is 11% higher. In Column 5, we use household consumption as the outcome and also find comparable, though generally smaller, effects; the smaller effects are likely due to richer households not consuming all that they earn. These results using income and consumption thus are consistent with the wage results, and suggest that sample selection bias does not seriously bias out wage analysis.

In the even columns of Table 4, we control for head's father's education as a further method to address concerns about ability bias. This does not change estimated returns to English ability much.

In Table 5, we allow the effect English-language skills to vary various observable characteristics. There is strong evidence of language-skill complementarity, with more experienced (see results by age), educated (by education) and able (by SSLC exam performance) workers

receiving a higher return to English-language skills. This has also been found in Israel by Berman, Lang and Siniver (2003).

5.1 Mechanism: Role of Occupation

In order to see if our results are driven by the occupational choices we run our specification with different occupational categories as defined by the National Classification of Occupations- 1968. We do not find differences in occupational choice to explain our estimated effects of English-language skills on wages.¹⁴

6 Conclusion

In this paper we have shown that there are large, statistically significant returns to English-language skills in India. We also find strong evidence of language-skill complementarity, with more experienced, educated and able workers are receiving significantly higher returns to English-language skills. We argue that estimating returns to English-language skills is not the same as returns to English-medium instruction.

Our results are particularly relevant for policy makers in developing countries where there has been constant debate whether to promote the local language or a more globally accepted language like English. This is especially true for India where different states are currently debating on what medium of instruction to be encouraged in the primary schooling. Our results show that language skills can have a far reaching impact on not just the treated but on household well-being as well. Roy (2007) investigates a policy experiment in West Bengal in India where teaching of English was abolished from primary schools. He does not find much beneficial effect of this policy on school enrollment, though there was significant increase in educational attainment, as measured by years of schooling.

¹⁴We like to emphasize that we view this result as a test for mechanism and not controlling for omitted variables.

Taken together, English-language skills form an important part of human capital. We quantify the returns to English-language skills not attempted previously due to lack of data in case of India. We also find evidence that the estimated returns are smaller than what has been estimated in case of US immigrants. We do not find occupation to be driving our results. It is a matter of future research to examine the effects of English-language skills on social networks and ties.

References

- Angrist, J., A. Chin and R. Godoy. "Is Spanish-Only Schooling Responsible for the Puerto Rican Language Gap?." *Journal of Development Economics*, February 2008, 85, pp. 105–108.
- Angrist, J. and V. Lavy. "The Effect of a Change in Language of Instruction on the Returns to Schooling in Morocco." *Journal of Labor Economics*, January 1997, 15 (1), pp. S48–76.
- Berman, E., K. Lang, and E. Siniver. "Language-Skill Complementarity: Returns to Immigrant Language Acquisition." *Labour Economics*, 2003, (10), pp. 265-290.
- Bleakley, H. and A. Chin. "Language Skills and Earnings: Evidence from Childhood Immigrants." *Review of Economics and Statistics*, May 2004, pp. 481–496.
- Card, D. "Chapter 30: The Causal Effect of Education on Earnings." *Handbook of Labor Economics, Volume 3*, in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*. Amsterdam: Elsevier Science, North-Holland, 1999, pp. 1801-1863.
- Clingingsmith, D. "Industrialization, Bilingualism, and Linguistic Heterogeneity in Mid-20th Century India." Case Western Reserve University unpublished paper, November 2008.
- He, F., L. Linden and M. MacLeod. "'How to Teach English in India: Testing the Relative Productivity of Instruction Methods within the Pratham English Language Education Program." Columbia University unpublished paper, July 2008.
- Desai, S, R. Vanneman, and National Council of Applied Economic Research, New Delhi . India Human Development Survey (IHDS), 2005 [Computer file]. ICPSR22626-v5. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2009-06-22.

- Hohenthal, A. "English in India: Loyalty and Attitudes." *Language in India*, May 2003, 3.
- Kapur, S. and T. Chakraborty "English Language Premium: Evidence from a Policy Experiment in India." Washington University in St. Louis unpublished paper, September 2008.
- Lang, K. and E. Siniver. "The Return to English in a Non-English Speaking Country: Russian Immigrants and Native Israelis in Israel." *NBER Working Paper 12464*, 2006, National Bureau of Economic Research, Cambridge, MA.
- Levinsohn, J. "Globalization and the Returns to Speaking English in South Africa," in Ann Harrison, ed., *Globalization and Poverty*. Chicago: University of Chicago Press, 2007, pp. 629-646.
- Munshi, K. and M. Rosenzweig. "Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy." *BREAD Working Paper 038*, 2003, Bureau for Research in Economic Analysis of Development.
- Munshi, K. and M. Rosenzweig. "Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy." *American Economic Review*, September 2006, 96 (4), pp. 1225–1252.
- Roy, J. "Can Changing Academic Standards Affect Educational Outcomes? Evidence from a Policy Experiment in India." Georgetown University unpublished paper, 2004.
- Shastri, G.K. "Human Capital Response to Globalization: Education and Information Technology in India." University of Virginia unpublished paper, October 2008.

TABLE 1
English-Speaking Ability in India

Variable Name	Converses fluently in English	Converses a little in English (but not fluent)	Total with English ability
All Individuals Aged 18-65	3.79 (0.19)	16.23 (0.36)	20.03 (0.40)
<i>By Sex:</i>			
Male	4.96 (0.21)	20.96 (0.40)	25.92 (0.43)
Female	2.62 (0.15)	11.53 (0.31)	14.16 (0.34)
<i>By Age:</i>			
18-35	4.35 (0.25)	20.21 (0.40)	24.56 (0.43)
36-50	3.25 (0.17)	12.80 (0.33)	16.05 (0.36)
51-65	2.95 (0.16)	9.70 (0.29)	12.65 (0.33)
<i>By Educational Attainment:</i>			
No Completed Schooling	0.00 (0.00)	0.03 (0.01)	0.03 (0.01)
Some Primary (1-4)	0.07 (0.02)	1.14 (0.10)	1.21 (0.10)
Primary (5-9)	0.47 (0.06)	10.33 (0.30)	10.81 (0.31)
Secondary (10-14)	5.98 (0.23)	49.52 (0.49)	55.50 (0.49)
Graduate (> 15)	35.34 (0.47)	53.20 (0.49)	88.54 (0.31)
<i>By Social Group:</i>			
Scheduled Tribes	2.53 (0.15)	8.37 (0.27)	10.91 (0.31)
Scheduled Castes	1.54 (0.12)	11.37 (0.31)	12.91 (0.33)
Other Backward Castes	2.93 (0.16)	14.69 (0.35)	17.63 (0.38)
Higher Castes	7.57 (0.26)	25.98 (0.43)	33.55 (0.47)

Table continues on next page.

TABLE 1
English-Speaking Ability in India (Continued)

Variable Name	Converses fluently in English	Converses a little in English (but not fluent)	Total with English ability
<i>By Geography:</i>			
Urban	9.00 (0.28)	25.83 (0.43)	34.84 (0.47)
Rural	1.68 (0.12)	12.37 (0.32)	14.06 (0.34)
East	2.29 (0.14)	10.07 (0.30)	12.36 (0.32)
Northeast	7.50 (0.26)	29.42 (0.45)	36.95 (0.48)
North	3.39 (0.18)	17.77 (0.38)	21.16 (0.40)
South	5.38 (0.22)	16.71 (0.37)	22.09 (0.41)
West	3.57 (0.18)	17.47 (0.37)	21.05 (0.40)

TABLE 1
English-Speaking Ability in India

Variable Name	Converses fluently in English	Converses a little in English (but not fluent)	Total with English ability
All Individuals Aged 18-65	3.79 (0.19)	16.23 (0.36)	20.03 (0.40)
<i>By Sex:</i>			
Male	4.96 (0.21)	20.96 (0.40)	25.92 (0.43)
Female	2.62 (0.15)	11.53 (0.31)	14.16 (0.34)
<i>By Age:</i>			
18-35	4.35 (0.25)	20.21 (0.40)	24.56 (0.43)
36-50	3.25 (0.17)	12.80 (0.33)	16.05 (0.36)
51-65	2.95 (0.16)	9.70 (0.29)	12.65 (0.33)
<i>By Educational Attainment:</i>			
No Completed Schooling	0.00 (0.00)	0.03 (0.01)	0.03 (0.01)
Some Primary (1-4)	0.07 (0.02)	1.14 (0.10)	1.21 (0.10)
Primary (5-9)	0.47 (0.06)	10.33 (0.30)	10.81 (0.31)
Secondary (10-14)	5.98 (0.23)	49.52 (0.49)	55.50 (0.49)
Graduate (> 15)	35.34 (0.47)	53.20 (0.49)	88.54 (0.31)
<i>By Social Group:</i>			
Scheduled Tribes	2.53 (0.15)	8.37 (0.27)	10.91 (0.31)
Scheduled Castes	1.54 (0.12)	11.37 (0.31)	12.91 (0.33)
Other Backward Castes	2.93 (0.16)	14.69 (0.35)	17.63 (0.38)
Higher Castes	7.57 (0.26)	25.98 (0.43)	33.55 (0.47)

Table continues on next page.

TABLE 1
English-Speaking Ability in India (Continued)

Variable Name	Converses fluently in English	Converses a little in English (but not fluent)	Total with English ability
<i>By Geography:</i>			
Urban	9.00 (0.28)	25.83 (0.43)	34.84 (0.47)
Rural	1.68 (0.12)	12.37 (0.32)	14.06 (0.34)
East	2.29 (0.14)	10.07 (0.30)	12.36 (0.32)
Northeast	7.50 (0.26)	29.42 (0.45)	36.95 (0.48)
North	3.39 (0.18)	17.77 (0.38)	21.16 (0.40)
South	5.38 (0.22)	16.71 (0.37)	22.09 (0.41)
West	3.57 (0.18)	17.47 (0.37)	21.05 (0.40)

Table 2. Summary Statistics

	All Individuals in Wage Employment (1)	All Male Household Heads (2)
Log hourly wage last year	2.07 (0.79)	2.31 (0.80)
Log household income last year		10.32 (0.98)
Log monthly per capita consumption last year		6.50 (0.68)
Fluent English	0.04 (0.20)	0.04 (0.20)
Little English	0.14 (0.34)	0.17 (0.37)
Male	0.71 (0.45)	1.00 (0.00)
Urban residence	0.26 (0.44)	0.29 (0.45)
Age	36.64 (11.61)	44.16 (11.00)
Social group: scheduled tribes	0.11 (0.31)	0.08 (0.28)
Social group: scheduled castes	0.28 (0.45)	0.22 (0.42)
Social group: other backward castes	0.34 (0.47)	0.35 (0.48)
Social group: Muslims	0.10 (0.29)	0.11 (0.32)
Years of schooling completed: 0	0.41 (0.49)	0.32 (0.46)
Years of schooling completed: 1	0.01 (0.07)	0.01 (0.08)
Years of schooling completed: 2	0.02 (0.15)	0.03 (0.16)
Years of schooling completed: 3	0.03 (0.16)	0.03 (0.18)
Years of schooling completed: 4	0.04 (0.19)	0.05 (0.22)
Years of schooling completed: 5 (Primary school completion)	0.08 (0.27)	0.09 (0.29)
Years of schooling completed: 6	0.03 (0.16)	0.03 (0.17)
Years of schooling completed: 7	0.05 (0.21)	0.05 (0.22)
Years of schooling completed: 8 (Middle school completion)	0.06 (0.23)	0.07 (0.25)
Years of schooling completed: 9	0.07 (0.25)	0.07 (0.26)
Years of schooling completed: 10 (Secondary school completion)	0.08 (0.28)	0.11 (0.31)
Years of schooling completed: 11	0.01 (0.12)	0.02 (0.13)
Years of schooling completed: 12 (Higher secondary school completion)	0.05 (0.21)	0.05 (0.23)
Years of schooling completed: 13	0.00 (0.04)	0.00 (0.05)
Years of schooling completed: 14	0.00 (0.05)	0.00 (0.05)
Years of schooling completed: 15 (Bachelor's degree or more)	0.07 (0.26)	0.08 (0.26)
Performance on SSLC exam: Class I	0.05 (0.22)	0.05 (0.21)
Performance on SSLC exam: Class II	0.11 (0.32)	0.14 (0.35)
Performance on SSLC exam: Class III	0.04 (0.19)	0.05 (0.21)
Has repeated or failed a grade	0.14 (0.35)	0.14 (0.35)
Has a permanent job	0.15 (0.35)	0.13 (0.33)
Has a government job	0.11 (0.31)	0.10 (0.29)
Has ever moved from current village, town or city	0.23 (0.42)	0.26 (0.44)
Number of observations	44650	33376

Notes: The sample consists of individuals aged 18-65 from the India Human Development Survey, 2005. Standard deviations are shown in parentheses below the means.

Table 3. Effect of English-Speaking Ability on Wages

	No Controls (1)	Demographic Controls (2)	Education Controls (3)	Geographic Controls (4)	'Ability' Controls (5)	District Fixed Effects (6)
Fluent English	1.489*** (0.0238)	1.271*** (0.0222)	0.609*** (0.0300)	0.397*** (0.0287)	0.340*** (0.0298)	0.321*** (0.0288)
Little English	0.823*** (0.0233)	0.655*** (0.0178)	0.229*** (0.0201)	0.142*** (0.0199)	0.129*** (0.0205)	0.120*** (0.0184)
Male		0.503*** (0.00970)	0.405*** (0.00889)	0.372*** (0.00831)	0.375*** (0.00823)	0.359*** (0.00800)
Age		0.0342*** (0.00243)	0.0350*** (0.00209)	0.0269*** (0.00190)	0.0268*** (0.00187)	0.0263*** (0.00184)
Age squared/100		-0.0322*** (0.00305)	-0.0301*** (0.00267)	-0.0220*** (0.00245)	-0.0219*** (0.00242)	-0.0215*** (0.00237)
Social group: scheduled castes		-0.315*** (0.0138)	-0.193*** (0.0127)	-0.111*** (0.0116)	-0.105*** (0.0114)	-0.0938*** (0.0113)
Social group: other backward castes		-0.304*** (0.0127)	-0.227*** (0.0124)	-0.120*** (0.0117)	-0.117*** (0.0117)	-0.0838*** (0.0115)
Social group: scheduled tribes		-0.447*** (0.0146)	-0.313*** (0.0146)	-0.178*** (0.0146)	-0.174*** (0.0143)	-0.124*** (0.0145)
Social group: Muslims		-0.230*** (0.0154)	-0.105*** (0.0148)	-0.165*** (0.0144)	-0.163*** (0.0143)	-0.122*** (0.0146)
Years of schooling completed: 1			0.0979** (0.0400)	0.0761** (0.0378)	0.0820** (0.0378)	0.0895** (0.0368)
Years of schooling completed: 2			0.0533** (0.0233)	0.0426** (0.0211)	0.0466** (0.0212)	0.0621*** (0.0196)
Years of schooling completed: 3			0.0587*** (0.0221)	0.0219 (0.0214)	0.0328 (0.0211)	0.0486*** (0.0187)
Years of schooling completed: 4			0.0698*** (0.0166)	0.0235 (0.0153)	0.0336** (0.0155)	0.0506*** (0.0147)
Years of schooling completed: 5 (Primary school completion)			0.180*** (0.0161)	0.116*** (0.0134)	0.128*** (0.0128)	0.115*** (0.0132)
Years of schooling completed: 6			0.222*** (0.0209)	0.120*** (0.0193)	0.132*** (0.0195)	0.136*** (0.0191)
Years of schooling completed: 7			0.247*** (0.0183)	0.166*** (0.0174)	0.181*** (0.0175)	0.171*** (0.0172)
Years of schooling completed: 8 (Middle school completion)			0.332*** (0.0170)	0.202*** (0.0159)	0.221*** (0.0164)	0.217*** (0.0158)
Years of schooling completed: 9			0.338*** (0.0166)	0.223*** (0.0157)	0.253*** (0.0165)	0.240*** (0.0156)
Years of schooling completed: 10 (Secondary school completion)			0.531*** (0.0186)	0.384*** (0.0197)	0.381*** (0.0575)	0.346*** (0.0482)
Years of schooling completed: 11			0.372*** (0.0373)	0.345*** (0.0346)	0.339*** (0.0605)	0.313*** (0.0550)
Years of schooling completed: 12 (Higher secondary school completion)			0.610*** (0.0257)	0.487*** (0.0249)	0.453*** (0.0563)	0.435*** (0.0504)
Years of schooling completed: 13			0.601*** (0.0914)	0.500*** (0.0838)	0.473*** (0.0947)	0.502*** (0.0936)
Years of schooling completed: 14			0.665*** (0.0940)	0.589*** (0.0961)	0.537*** (0.109)	0.516*** (0.101)
Years of schooling completed: 15 (Bachelor's degree or more)			0.960*** (0.0276)	0.825*** (0.0264)	0.745*** (0.0570)	0.713*** (0.0506)
Performance on SSLC exam: Class I					0.234*** (0.0640)	0.239*** (0.0534)
Performance on SSLC exam: Class II					0.0626 (0.0605)	0.0665 (0.0513)
Performance on SSLC exam: Class III					-0.0999 (0.0613)	-0.0732 (0.0517)
Has repeated or failed a grade					-0.0687*** (0.0142)	-0.0660*** (0.0118)
Urban residence				0.413*** (0.00963)	0.405*** (0.00949)	0.341*** (0.0117)
State fixed effects	No	No	No	Yes	Yes	No
District fixed effects	No	No	No	No	No	Yes
R-squared	0.242	0.378	0.431	0.514	0.518	0.569
Number of observations	44650	44650	44650	44650	44650	44650

Notes: The sample consists of individuals aged 18-65 who reported wage and salary work from the India Human Development Survey, 2005. The dependent variable is log hourly wage. The omitted social group are the high castes (88%), Christians, Sikhs and Jains. The omitted educational attainment group is zero years of schooling completed. Robust standard errors are shown in parentheses. Asterisks denote significance levels (*=.10, **=.05, ***=.01).

Table 4. Effects of English-Speaking Ability for Male-Headed Households

	Dependent Variable:					
	Household Head's Log Hourly Wage		Household's Log Total Income Last Year		Household's Log Monthly Per Capita Expenditure	
	District Fixed Effects (1)	Add Father's Education (2)	District Fixed Effects (3)	Add Father's Education (4)	District Fixed Effects (5)	Add Father's Education (6)
Fluent English	0.330*** (0.0378)	0.312*** (0.0380)	0.337*** (0.0365)	0.320*** (0.0364)	0.257*** (0.0231)	0.235*** (0.0232)
Little English	0.163*** (0.0238)	0.160*** (0.0237)	0.113*** (0.0240)	0.106*** (0.0239)	0.103*** (0.0140)	0.0980*** (0.0142)
R-squared	0.556	0.557	0.369	0.373	0.468	0.473
Number of observations	21057	20672	32916	32272	33340	32688

Notes: The sample consists of male heads of household aged 18-65 from the India Human Development Survey, 2005. Each column reports the results of a separate regression that also controls for an age quadratic, social group dummies, education dummies, urban dummy, district dummies, exam performance dummies, and dummy for failing or repeating a grade (i.e., the same specification as in Table 3, Column 6). The even columns also control for a full set of dummies for father's years of schooling completed. Robust standard errors are shown in parentheses. Asterisks denote significance levels (*=.10, **=.05, ***=.01).

Table 5. Heterogeneity in Effect of English-Speaking Ability on Log Hourly Wages

Variable 1: Variable 2: Variable 3: <i>Omitted group:</i>		Male	Urban	age 18-35 age 36-50	ST SC OBC	BA Secondary	Exam Div. I Exam Div. II	Permanent Job	Government Job	Mover
	<i>Female</i>	<i>Rural</i>	<i>age 51-65</i>	<i>Higher Castes</i>	<i>< 10 yrs schooling</i>	<i>No SSLC or Div. III</i>	<i>Casual Job</i>	<i>Private Sector Job</i>	<i>Non-mover</i>	
Fluent English	0.321*** (0.0288)	0.364*** (0.0555)	0.292*** (0.0507)	0.703*** (0.0477)	0.381*** (0.0323)	0.203* (0.109)	0.233*** (0.0590)	0.139*** (0.0311)	0.268*** (0.0315)	0.306*** (0.0363)
Little English	0.120*** (0.0184)	0.156*** (0.0460)	0.107*** (0.0251)	0.590*** (0.0478)	0.168*** (0.0220)	0.0513* (0.0284)	0.133*** (0.0238)	-0.00875 (0.0164)	0.0379* (0.0203)	0.0959*** (0.0224)
Fluent English x Variable 1		-0.0495 (0.0536)	0.0510 (0.0517)	-0.593*** (0.0506)	-0.0617 (0.0904)	0.175 (0.121)	0.245*** (0.0778)	0.197*** (0.0373)	-0.0412 (0.0430)	0.0458 (0.0388)
Little English x Variable 1		-0.0404 (0.0460)	0.0331 (0.0272)	-0.671*** (0.0492)	0.0183 (0.0590)	0.137** (0.0603)	0.109* (0.0579)	0.223*** (0.0282)	0.107*** (0.0354)	0.0754*** (0.0258)
Fluent English x Variable 2				-0.251*** (0.0512)	-0.198*** (0.0676)	0.152 (0.116)	0.0387 (0.0678)			
Little English x Variable 2				-0.300*** (0.0501)	-0.147*** (0.0324)	0.103** (0.0413)	-0.0560* (0.0337)			
Fluent English x Variable 3					-0.0771* (0.0468)					
Little English x Variable 3					-0.0450 (0.0345)					
Has Permanent Job							0.611*** (0.0203)			
Has Government Job								0.714*** (0.0293)		
Has Ever Moved										0.0692*** (0.0106)

Notes: The sample consists of individuals aged 18-65 who reported wage and salary work from the India Human Development Survey, 2005 (44,650 observations).

The dependent variable is log hourly wage. Each column reports the results of a separate regression that also controls for an age quadratic, social group dummies, education dummies, urban dummy, district dummies, exam performance dummies, and dummy for failing or repeating a grade (i.e., the same specification as in Table 2, Column 3). Robust standard errors are shown in parentheses.

Asterisks denote significance levels (*=.10, **=.05, ***=.01).

