The Returns to English-Language Skills in India

Mehtabul Azam   Aimee Chin

World Bank & IZA[*]   University of Houston & NBER[†]

Nishith Prakash

University of Connecticut, IZA, INSIDE & CReAM[‡]

April 2011[§]

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. In this paper, we use individual-level data from the India Human Development Survey, 2005 to quantify the effects of English-language skills on wages. After controlling for age, social group, schooling, geography and proxies for ability, we find that hourly wages are on average 34% higher for men who speak fluent English and 13% higher for men who speak a little English relative to men who do not speak English. The return to fluent English is as large as the return to completing secondary school and half as large as the return to completing a Bachelor's degree. Additionally, we find that more experienced and more educated workers receive higher returns to English. The complementarity between English skills and education appears to have strengthened over time–only the more educated among young workers receive a premium for English-speaking ability, whereas older workers across all education groups do.

JEL Classification: J31, J24, O15
Keywords: English Language, Human Capital, India

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 a microdata set containing measures of both earnings and English ability.[1] We take advantage of a recently available nationally representative individual-level data set, the India Human Development Survey (IHDS), 2005, to provide the first estimates of the returns to English-language skills in India. A secondary 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 some 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, age and sex.

A major challenge to estimating the returns to English is the likely endogeneity of language skills in the earnings equation. A priori, omitted variables bias seems to be a serious concern; for example, omitted ability or local labor market conditions can be correlated with both language skills and earnings. We exploit the 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 in the Indian context. As well, the large sample size and detailed geographic identifiers permit controlling for labor market characteristics via district fixed effects. Nonetheless, even after controlling for more detailed variables than is typically available in household surveys, there could still be bias in our estimates of the effect of English-language skills. Unobserved ability could still be a confounding factor to the extent that our included variables capture it inadequately, and additionally there could be biases due to measurement error and reverse

---

[1]Munshi and Rosenzweig (2006) and Chakraborty and Kapur (2008) estimate the returns to attending a school using English as the medium of instruction. We explain in subsection 2.2 that this is not the same as the returns to English-language skills.

causality (Section 3 discusses this further); without an experiment that randomly assigns English skills or a valid instrumental variable for English skills, it is difficult to convincingly overcome these remaining identification issues. However, we believe that even if our estimated returns to English do not have a causal interpretation, they are still useful. On the one hand, very little information exists on the relationship between English skill and earnings in developing countries, so even if our estimates were purely descriptive they would still add to knowledge. On the other hand, English skills are not the only type of human capital that individuals and policymakers can invest in, and a comparison of the returns to English to returns to other types of human capital (such as schooling or job training) is revealing about the relative value of English even if the estimated returns do not have a causal interpretation. This is because estimating returns to these other types of human capital faces the same identification issues as estimating the returns to English, so comparisons of the returns might still be meaningful.

Our main findings are as follows. First, English-language skills are strongly positively associated with earnings. After controlling for age, social group, schooling, geography and proxies for ability and geography, we find that hourly wages are on average 34% higher for men who speak fluent English and 13% higher for men who speak a little English relative to men who speak no English. These estimates are not only statistically significant, they are also economically significant. For example, the return to fluent English is as large as the return to completing secondary school and half as large as the return to completing a Bachelor's degree. Second, there is considerable heterogeneity in the returns to English. More experienced and more educated workers receive higher returns to English. The complementarity between English skills and education appears to have strengthened over time–only the more educated among young workers earn a premium for English skill, whereas older workers across all education groups do.

This study is of interest for several reasons. Foremost, knowing the returns to English would help individuals and policymakers in India make decisions about how much to invest in

English skills. Language skills are costly to acquire, and it is difficult to make optimal choices without knowledge about the expected benefits of English-language skills. Additionally, this study informs on the more general question of the value of English in a context where English is not a prevalent language. English is often used as a lingua franca–the language of communication among two parties who do not share a common native language–and many countries, even ones that are not former British or American colonies, invest in English skills.

The remainder of this paper is organized as follows. Section 2 provides some 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 concludes.

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

---

[2]Under 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.

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.[3] 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.[4] A common language 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 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

---

[3]In 2001, 40% of the population named Hindi as their mother tongue; the next language with most native speakers, Telugu, claimed only 8%.

[4]See Shastry (2011) for more on within-district variation in languages spoken.

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, comprised of 4% who can converse fluently in English and 16% who 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 probably largely due to 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 degree can speak English as compared to 56% for those who have completed secondary schooling (10-14 years of schooling completed), 11% for those who have completed 5-9 years, and virtually nil for 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.[5] According to the 1986 All-India Education Survey, 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

---

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

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, and exams may be written in English, Hindi or the state language.

Next, we examine percent speaking English by social group. In India, the two most disadvantaged social groups are the schedule tribes (STs) and scheduled castes (SCs).[6] The Other Backward Classes (OBCs) are above the SCs in ritual standing, but are also much worse off than the high castes. English ability is greater among members of higher castes than members of lower castes or the scheduled tribes. This is likely related to the lower educational attainment among members of disadvantaged social groups, and in the case of the scheduled tribes, of their geographic isolation.

There is considerable geographic variation in the prevalence of English in India, as can be seen in Figure 1, which shows the mean English ability by state. It is beyond the scope of this paper to account for all these cross-state differences, however we do describe English ability along several geographic dimensions on the bottom of Table 1. First, there is a large difference in English ability by urban status: 35% of individuals living in urban areas report to have ability to speak English as compared to only 14% living in rural areas. Second, individuals living in districts with greater historical prevalence of English skill are more likely to speak English today: 26% of individuals living in districts with above-median share of the 1961 population speak English, compared to 14% of those living in other districts. Additionally, English-speaking ability is more widespread in districts that had greater linguistic diversity in 1961, or that had an information technology firm in 2003.

---

[6]STs are distinguished by their tribal culture and physical isolation. SCs are groups with low social and ritual standing. In the British era, these were called the depressed classes, and colloquially they have also been called the untouchables and backward classes though these terms are out of favor. The Constitution (Scheduled Castes) Order of 1950 and the Constitution (Scheduled Tribes) Order of 1950 lists which castes, races and tribes are designated SCs and STs, respectively.

Having described 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), where the latter is an unpublished manuscript. Both estimate the returns to attending a school with English (as opposed to some native language) as the medium of instruction. Munshi and Rosenzweig collected their own data on Maharashtrians living in Dadar, which is located 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.[7]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.[8]

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.[9] Angrist and Lavy (1997), for example, find that French-language skills significantly

---

[7]These returns are described in greater detail in Munshi and Rosenzweig (2003).

[8]Estimates controlling for secular cohort trends suggest somewhat larger effects.

[9]That is, we might think of the returns to English-medium schooling as a reduced-form relationship

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 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.[10] A second difference is that our study uses a large, nationally representative data set, which enables us to explore potential heterogeneity in returns to English-language skills along various dimensions (below, we will allow returns to vary by sex, age, education, social group and geographic variables). Munshi and Rosenzweig's findings come from one community in Mumbai, and Chakraborty and Kapur's findings come from a policy change in one state, 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

---

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 that changed the medium of instruction from English to Bengali increased educational attainment.

[10]They 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.

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.

This study makes several contributions to the existing literature. 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, 2005 helps us clear this hurdle. Additionally, it adds to the small 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 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.[11]

Among the benefits of having English-language skills is getting a better job. Though it is

---

[11]Roy (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.

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.[12] This skill price is determined by the supply of and demand for English-proficient workers. The recent availability of IHDS data, 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:

$$y_i = \alpha + \beta English_i + \pi X_i + e_i \tag{1}$$

where y_i is the log hourly 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.

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, 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

---

[12]As 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. Moreover, it is of interest to explore heterogeneity in returns to English, which we can do because of our large, diverse sample.

proficient) labor force in a particular region attract firms to locate there. Shastry (2010) 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 (2007) 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 our preferred specification, we control for geographic conditions more finely by including district of residence fixed effects, i.e., we are comparing individuals who speak English and those who do 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 may be 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. A SSLC makes one eligible for further schooling, and a better SSLC exam performance enables one to attend better schools. This exam is typically taken at the end of 10th grade, and the passing categories, from highest to lowest level of distinction, are Class

I, II and III.[13] Performance on the SSLC exam–even if measured in only a few categories–seems like a credible proxy for ability in the Indian context, and is akin to controlling for aptitude test scores to address the ability bias when estimating the returns to schooling in the U.S. context. 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 30% of individuals aged 18-65 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 for ability 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:

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

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 $Ability_i$ are proxies 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, measurement error and reverse causality. First, of course it is impossible to rule out that we have left out some relevant variable. However, it should be noted that in order to change our results materially, it would have to be a variable that, conditional on all the explanatory variables mentioned above, has a strong relationship with wages or English ability. Below, we find that although the proxies for ability are significant predictors of wages, nevertheless their inclusion does not change

[13]In the IHDS, among individuals aged 18 to 65 who have completed 10th grade or higher, 20% reported Class I, 53% reported Class II, 17% reported Class III, and 9% had a missing value. Below, we will form dummies for each of these four categories, and use the latter as the omitted reference category. We observe that people in the Class II, Class III and missing value categories have similar wages (holding other explanatory variables constant), while people in the Class I category have significantly higher wages. Thus it appears that the people not reporting SSLC performance are primarily Class II and III performers.

the coefficients for English skill much, which suggests that selection on unobserved variables is unlikely to account for much of our estimated returns to English (Altonji, Elder and Taber (2005)). Second, measurement error is potentially a serious problem because the English-speaking ability variable available from the IHDS is self-reported (rather than based on some objective test) and constrained to a few categories (speaks fluent English, a little English or no English). Bleakley and Chin (2004) and Dustmann and van Soest (2002), using U.S. Census data and German Socio-Economic Panel data which have similar language measures, find that returns to language skill are higher after correcting for measurement error. If the nature of measurement error in the English ability measure from the IHDS is similar, then we might expect our coefficients for English ability to be downward biased. While we know classical measurement error leads to attenuation bias, the overall bias in our estimates due to measurement error is unclear because the direction and extent of non-classical measurement error is not known.[14] Third, reverse causality could be a source of bias. Being in a higher-paying job might cause workers to develop better English-language-skills, such as through being able to afford English lessons or getting more exposure to English in the workplace. If so, then the estimated coefficient for English skills would exceed the true returns to English because it encapsulates the reverse effect too. For these reasons, our empirical strategy may not yield estimates of the returns to English that have a causal interpretation. Bleakley and Chin use an instrumental variables strategy based on U.S. immigrants' age at arrival and country of origin to obtain the causal returns to English, and find that the instrumental variable estimate (which avoids both the downward bias from classical measurement error and the upward bias from omitted "ability" and reverse causality) exceeds the ordinary least squares estimate (which is subject to all those biases), which suggests that attenuation bias

---

[14]Bleakley and Chin use data from the National Adult Literacy Survey to analyze the relationship between the self-assessed categorical measure of English skill to a better measure based on a literacy test, and quantify bias from measurement error. In their case, it turns out there is some upward bias due to non-classical measurement error (in which the measurement error is correlated with the English skill variable), but the downward bias due to classical measurement error is larger by far. In our case, without the benefit of an external data set containing our measure of English skill and a better one for India, we cannot make more definitive statements about measurement error.

dominates other sources of bias in their setting. We do not know, however, whether the relative size of the various biases found in their setting applies to ours.

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.[15] 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.[16]

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 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 alternative 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 various subsamples used below: men in

---

[15]The 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. It is a stratified sample, and for our data analysis we use the IHDS design weights ("sweight") to obtain nationally representative statistics.

[16]The 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.

wage employment, male household heads in wage employment, all male household heads, and women in wage employment. We discuss our dependent variables in greater detail here. The top row reports that mean log hourly wages is 2.25 for male wage workers, 2.31 for the subset that is a household head, and 1.62 for female wage workers. The specific measure of wages that we use is hourly wage for primary job.[17] The second row reports that mean log household income last year is 10.3. This income measure captures all sources of income from all members of the household.[18] The third row reports the mean log monthly consumption per capita. This measure is formed based on measures of the household's consumption from forty-seven categories of goods and services.[19]

5   The Returns to English-Language Skills

5.1   Main Results

Using the sample of male wage earners aged 18 to 65, we estimate Equation 2 using OLS.[20] These results are shown in Table 3, with each column adding more controls. Column 1 presents the raw difference in log hourly wage by English proficiency. Compared to log wages of 2.07 for men who have no English ability, men who speak a little English earn 0.69

---

[17]IHDS asked respondents about pay for each job held last year, and we use data on the first job. 93% of wage earners reported only one job, so not surprisingly the findings are the same when we use total wages across jobs instead of wages from primary job.

[18]The IHDS is unusual among Indian household data sets for its rich measures of income. Households were asked about over fifty components of income and our measure of household income takes the sum across all sources. By far the largest sources were wages and salaries, farm net income and non-farm business net income, but also included are such things as government transfers, pension income and income from investments. Except for wages and salary, all the income components are collected at the level of the household, which is why we use household income rather than individual income.

[19]Thirty of the consumption categories are frequently purchased items such as food (including home-produced food), housing and transportation, and use a 30-day time frame. The remaining seventeen categories are less frequently purchased items such as school fees and durables, and use a 365-day time frame. The consumption module of the IHDS is based on the short-form consumption questions in the National Sample Survey (NSS). The NSS is a nationwide, large-scale, multi-purpose household survey administered by the Government of India's Ministry of Statistics and Programme Implementation, and the Consumption Expenditure Survey has been part of the NSS quinquennially since 1972-73.

[20]We estimate models separately for men and women. We discuss results for women in subsection 5.3.

log points more, and men who speak English fluently earn 1.37 log points more.[21] These differences are large, but likely overstate the true returns to English; for example, in Table 1, we observed 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 3, we control for 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.[22] 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 602 districts) and this does not make much difference. We regard the Column 6 specification–with the most comprehensive controls–as our preferred specification. The estimates in Column 6 suggest that, compared to men who have no English ability, men who are fluent in English have 34% higher hourly wages, and men who speak a little English have 13% higher hourly wages. These results

---

[21]In rupee terms, hourly wages are 10, 23 and 42 for men with no, little and fluent English.

[22]It is comforting that the estimated coefficients for English skill do not change much with the addition of these variables, and suggests that the estimated coefficients for the language variables is really due to language rather than remaining omitted variables. Had ability bias been important, then the inclusion of the proxies for ability should have changed the estimated coefficients for English skill more dramatically. We might therefore reason, along the lines of Altonji, Elder and Taber (2005), that bias from remaining unobservables is unlikely to be able to account for the estimated coefficients for English skill. As an aside, it is surprising that the adjusted R-squared does not increase much going from Column 4 to 5 despite the statistical and economic significance of the dummies for SSLC exam performance and grade repetition/failure. This appears to be due these variables' high degree of correlation with the educational attainment variables, and suggests that in wage regressions that do not control for ability, the estimated return to education in part reflects the effects of ability.

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.[23]

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. In particular, we may be concerned that English proficiency affects both participation in wage employment and wages such that our estimates may not give the unconditional effect of English proficiency on wages. In Appendix Table 1, we examine the selection into wage employment explicitly. Specifically, we estimate Equation 2 with probit, where the dependent variable is a dummy for being in wage employment and the sample is all men aged 18-65.[24] We find that men who speak a little English are not more or less likely to participate in wage employment relative to men who speak no English, suggesting that our coefficients for "little English" in the wage regression are not biased by selective participation. On the other hand, men who speak fluent English are on average 7.6 percentage points more likely to participate in wage employment relative to men who speak no English, so selection bias could be an issue. We address this issue by looking at household income and household consumption as outcomes in place of wages. These measures are available for everyone regardless of one's participation in wage employment. Our analysis with the income and consumption measures will be performed using male heads of households to avoid counting the same household multiple times. Just as in the sample of all men, in the subsample of men who are heads, those who speak English fluently are significantly more likely to participate in wage employment (Appendix Table 1, Column 2). However, in this same subsample, English skill is not a significant determinant of participation in employment (any type, including wage employment).[25] Thus, among male

---

[23]Note that these comparisons between returns to English and returns to schooling are made based on the Column 6 specification, so are conditional on SSLC exam performance among other things.

[24]An individual is coded as being in wage employment if he/she reports positive wages last year. We use the same specification as in Table 3, Column 6.

[25]An individual is coded as engaged in employment if he/she reports positive hours of work in wage employment, the family farm or non-farm family businesses.

heads, there is no evidence of selection into employment by English skill, and the measures of household income and consumption reflect at least in part male heads' labor input.[26]

In Table 4, we present the estimation results using household income and consumption as outcomes. In Column 1, we estimate the same model as in Table 3, Column 6 to demonstrate that very similar effects of English skill on hourly wages are found when we restrict analysis to 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 could be due to richer households saving more. These results using income and consumption thus are consistent with the wage results, and suggest that sample selection bias does not seriously bias our 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.[27] This does not change the estimated returns to English ability much, which is comforting along the lines we argued above when we included the controls for SSLC exam performance and grade repetition/failure; the father's education dummies are jointly significant predictors of male heads' wages, household income and household consumption.

5.2   Heterogeneity in Returns to English

Thus far, we have focused on estimating the average returns to English. In this subsection, we explore heterogeneity in returns to English by demographic characteristics and by geography.

By Demographic Variables: Table 5 reports the returns to English by age, education and

---

[26]As we mentioned in Section 4, income and consumption are measured only at the household level, not at the individual level. We can view household-level income and consumption either as noisy measures of their individual-level counterparts or as per se interesting outcomes.

[27]Data on father's education are collected only for male household heads so specifications controlling for this variable cannot be estimated using the broader sample.

social group. First, we consider age/cohort; given that we have a single cross-section, we cannot distinguish between age and cohort effects. We find that older cohorts receive significantly higher returns for English skills (Column 2). Among wage workers in the youngest age category, 18-35, those who speak English fluently earn 11% higher wages relative to those with no English ability, and those who speak a little English actually earn less (rows 1 and 2, respectively). Men aged 36-50 have returns of 49% for fluent English and 30% for a little English (add rows 1 and 3, and rows 2 and 4, respectively). Men aged 51-65 have returns of 68% for fluent English and 54% for a little English. Although demand for workers with English skill has expanded in recent decades due to the growth of trade and outsourcing, the supply of workers with such skill must also have grown rapidly lest we would have observed a rising English skill price. Of course the question arises as to why the skill price is not equalized across the cohorts, i.e., with the influx of young workers speaking English into the labor market, why aren't the wages of older workers bid down? Consistent with these empirical observations is a story where there are language-skill complementarities. For example, English skill enables workers to enter more lucrative career tracks, i.e., these are jobs that have better promotion (or career progress) and raise opportunities. During his tenure, an older vintage worker would have acquired significant work experience. Although a younger worker possesses the English skills that would have been necessary to obtain the entry-level position decades ago, he lacks the work experience, making him unable to substitute for the older vintage worker today.

Second, we examine the returns to English by educational attainment. There is suggestive evidence of higher returns to English for the more educated (Column 3). Among men who have not completed their secondary schooling (i.e., those with 0-9 years of completed schooling), those who are fluent in English have 24% higher wages relative to those with no English ability, and those who speak a little English have 7% higher wages. The positive coefficients for the interactions between English skill and being higher educated (rows 3-6) suggest that the return increases with education, however only one of these is statistically

significant at conventional levels.

Next, we allow the returns to vary by both cohort and education. Understandably, the results are less precise given the number of interaction terms in the model. To make the model more tractable, we just use two age categories instead of three: young (age 18-35) and older (age 36-65). The results are summarized in Figure 2 (which is derived from the coefficients reported in Table 5, Column 4). Perhaps the most striking feature of Figure 2 is that the education profile for returns to English is upward-sloping for younger men but basically flat for older men. Men aged 18-35 who have not completed secondary school do not receive any premium for their English skill. More educated young men, though, do earn significant returns to English: those who complete secondary school have a 26% return for speaking fluent English and a 5% return for speaking a little English, and those who complete a Bachelor's degree earn a 40% return and a 17% return, respectively. For older cohorts, even low educated men aged 36-65 receive a sizable English premium–it is 43% for fluent English, and 22% for a little English (which are not so different from the premia for college-educated young men!). In Column 5, we present the results of estimating a less demanding model that combines the two lower education groups, leaving us with only two education categories: Bachelor's degree and less. We likewise see that returns to English are higher for older workers, and educated young workers.

These results by age-education in Column 4 and 5 help put the earlier results by age alone and education alone into greater perspective. Earlier, we found larger returns to English for the older cohorts, with the youngest cohort earning only modest returns (Column 2). We found in Column 3 that the return to English was increasing in educational attainment, though often the interaction effects were weak. It is possible, though, that some of the age interactions are due to differences in education by age and vice versa. Columns 4 and 5 address this possibility, and reveal that returns to English are higher for older cohorts, and the subset of the young who are more educated.

Berman, Lang and Siniver (2003) and Lang and Siniver (2006) find evidence of language-

skill complementarity in the Israeli context. Immigrant workers in high-skilled occupations received larger wage increases when they learned Hebrew and English (on top of their own native language, Russian). In India, we also find evidence of language-skill complementarity. On the one hand, more experienced workers receive a higher return for English skill; this is based on the observation that older workers have a higher English premium than younger workers. On the other hand, for recent cohorts, more educated workers receive a higher return for English skill. It is interesting that such a language-education complementarity was not found for older cohorts. This is consistent with increasing complementarity between education and English skills over time. For example, at the entry level, workers with English skill may have been able to find a good job decades ago but now only the subset with more education would find a good job. This could be because it has become more competitive to get good jobs (because the supply of English-proficient workers has expanded so much), or because there are new jobs that require both higher education as well as English skill to perform.

A final demographic variable we consider is social group. We find that the returns to English are significantly lower for SCs relative to high castes, and no different for STs and OBCs relative to high castes.[28] The negative estimated coefficients for the interactions between SC and English skill in Column 6 in part reflect the lower education of SCs relative to the high castes, because once we allow for age-education-specific effects of English skill in Column 7, their magnitude decreases. However, a significant negative coefficient remains for the interaction between SC and fluent English; the point estimate suggests that STs who are fluent in English earn a 14% percentage point lower return.

By Geographic Variables: In Table 6, we allow the returns to English to vary by urban residence, district historical English prevalence, district historical linguistic heterogeneity and district IT presence.[29] We view the coefficients for these interactions as descriptive, as

---

[28]The interactions with STs are imprecisely estimated, though, because STs are only a small share (about 10%) of the population and therefore our sample.

[29]Note that the main effects of these variables are controlled for in our regressions via the urban dummy

these geographic variables are correlated with numerous variables along which the returns to English might differ.

First, we consider urban/rural residence. To the extent that jobs rewarding English tend to be located in cities (where higher levels of government, multinational firms, or information technology firms tend to be located), all else equal, we might expect higher returns to English in urban areas. However, urban areas also tend to have a larger supply of English-proficient workers (both because urban schools produce more English speakers than rural schools, and because cities may attract English speakers from elsewhere), so it is an empirical question whether urban areas do have higher returns to English. In Table 6, Column 2, we find no evidence of a differential return to English by urban/rural residence for men.

Second, we allow the returns to English to vary by historical English prevalence. We measure this as the share (from 0 to 1) of the population reporting English as a mother tongue or a second language in the individual's district of residence according to the 1961 Census.[30] Although districts with more widespread English in 1961 continue to have more widespread English today (as noted in subsection 2.1), they have significantly lower estimated returns to English (Table 6, Column 3).[31]This is consistent with the returns to English being lower in places where the supply of English-proficient workers is higher.

Third, we allow the returns to English to vary by historical linguistic heterogeneity. We use the measure in Clingingsmith (2007), which is one minus the Herfindahl index of mother tongues in the individual's district, where each language's share of the population to form the Herfindahl index is from the 1961 Census.[32] Thus, it can vary from 0 to 1, with higher values indicating less concentration (i.e., greater diversity of language in one's district).

---

and district fixed effects.

[30]We are very grateful to David Clingingsmith for sharing the 1961 Census district-level language data with us. In our sample of male wage earners, the English prevalence measure has a mean of 0.028 (s.d. of 0.032) and varies from 0.0006 to 0.234.

[31]Note we have demeaned the variables, so the coefficients for "Fluent English" and "Little English" reflect the returns computed at the average English prevalence.

[32]We are very grateful to David Clingingsmith for sharing the district linguistic heterogeneity data with us. In our sample of male wage earners, linguistic heterogeneity has a mean of 0.34 (s.d. of 0.21) and varies from 0.007 to 0.998.

Although districts that were more linguistically diverse in 1961 have more widespread English today, they have lower estimated returns to fluent English (Table 6, Column 4).[33] But this differential return disappears once we also control for historical English prevalence (Columns 6 and 7). Thus, while English skill may well be especially valuable in linguistically diverse places, these places tend to have more English speakers (in part because English is more likely to be taught in schools), such that their price of English skill is not higher.

Finally, we look at heterogeneity in returns by IT presence. We measure this as a dummy variable indicating whether the individual's district had any IT firm headquarters or branch according to the 2003 National Association of Software and Service Companies (NASSCOM) directory.[34] Though districts with an IT firm have more widespread English, they have significantly lower estimated returns to fluent English (Table 6, Column 5). The IT interaction effects decrease in magnitude with the addition of interactions of English skill with the other geographic variables and age-education (Columns 6 and 7). These results are consistent with IT firms choosing to locate in districts where English-proficient and educated workers are in greater supply, as Shastry (2011) has found.

There is overlap in the four geographic variables, and in Column 6 we add their interactions with English skill all at once. In Column 7, we also control for the age-education-specific effects of English skill. We continue to find significant negative coefficients for the interactions between district English prevalence and individual English ability. The point estimates imply that a district with a one percentage point higher share of the population speaking English in 1961 has a 1.4 percentage point lower return to English (fluent or a little). These results are consistent with an important role for the supply of English-proficient speakers. The growth in the supply of English-proficient workers has been especially fast in recent decades, and might be a factor behind the more limited returns that we estimated for young

---

[33]Again, we have demeaned the variables, so the coefficients for "Fluent English" and "Little English" reflect the returns computed at the average linguistic heterogeneity.

[34]We are very grateful to Kartini Shastry for sharing these data with us. In our sample of male wage earners, mean IT presence is 0.25 (s.d. of 0.43). We have tried different measures of IT presence, such as having at least a certain number of IT firms or IT employees, and results are similar.

workers (aged 18-35) relative to older workers.

5.3   Results for Women

Thus far, we have focused on the returns to English for men. While we are as interested in quantifying the returns to English for women too, two reasons led us to focus on men. First, selective participation in wage employment is a more serious problem for women. Only 22% of women work in a wage-and-salary job, compared to 54% of men. Participation in wage employment is more sensitive to English ability for women than men: Appendix Table 2, Column 4 suggests that women who are fluent in English are 17 percentage points more likely to be in wage employment relative to women with no English ability, and women who speak a little English are 4 percentage points more likely. Second, we do not have a convincing methodology for fixing the sample selection bias for women. For men, we addressed it by examining household-level income and consumption as outcomes in place of wages. While the analysis for men suggests that sample selection bias is minimal, this result may not necessarily carry over to women. Therefore, in this subsection, we discuss results for women with the caveat that bias from selective participation in wage employment could be present (along with the other biases mentioned in Section 3).

Table 7 presents select estimation results for female wage earners.[35] In Column 1, after controlling for age, social group, schooling, urban dummy, district fixed effects and proxies for ability, we find that women who are fluent in English earn 22% more relative to women with no English ability, and women who speak a little English earn 10% more.[36] In Column 2, we find evidence of skill-language complementarity. Older and more educated women have higher returns to English. Among young women with English skill, only those who are educated earn a positive return for their English skill, so the complementarity between

---

[35] More detailed tables of estimation results for women, paralleling Tables 3, 5 and 6 for men, are available from the authors.

[36] Though the point estimates of the returns to English are lower for women than men, it should be noted that these estimates are imprecise. Based on pooled regressions in which we allow returns to English to vary by sex, we cannot reject that the returns for women and men are equal.

English skills and education appears to have strengthened over time for women. In Column 3, we find that lower castes receive lower returns to English even after allowing for returns to English to differ by age-education. In Column 4, we find that historical English prevalence reduces the returns to English.

The most striking observation from Table 7 is how qualitatively similar the men's and women's results are. One difference worth noting, though, is that the urban dummy, historical linguistic heterogeneity,and IT presence variables tend to have positive interaction effects with English skill for women, but negative ones for men. For example, while returns to English do not vary by urban/rural residence for men, they do for women. Indeed, we find that on average women with English ability do not earn more than women with no English ability in rural areas, but they do in urban areas (in fact, the point estimate of urban women's returns to English are at least as large as men's). We leave a closer examination of these gender differences in returns to English to future work.

5.4   Role of Occupation and Industry

One mechanism through which English skill might affect wages is through occupational choice. Due to India's colonial past, many jobs in government and education require English. Also, due to the way science, engineering and other technical fields are typically taught in universities (in English) and due to the growth of international trade and outsourcing, many modern, technical jobs also require English to enter. But these jobs–which include government officials, teachers, engineers, physicians, and managers–tend to be attractive relative to other jobs in the economy. Thus, it is of interest to ask how much of the estimated returns to English is accounted for by occupational choice alone.

We decompose the overall returns to English into within- and between-occupation components by augmenting our base model with detailed occupation dummy variables and comparing these returns to the returns from the base model.[37] We find that for men, after

---

[37]For wage workers, the IHDS codes occupation into approximately 90 categories.

controlling for occupation dummies, the return to fluent English decreases to 22% and the return to a little English decreases to 7% (Appendix Table 2, Column 2). That is, even among wage workers *within* the same occupation, there is significantly higher pay for those with English skill. According to the point estimates in Column 2, occupation alone accounts for about one-third of the total estimated coefficient (from the base model in Column 1) for "fluent English" and about half of that for "little English". Thus, a majority of the overall returns to English is left unexplained, suggesting that the English premium is not merely an occupation premium. In Column 3, we find that industry choice plays a smaller role in the returns to English than occupational choice.[38] In Column 4, we control for both occupation and industry and still find a larger within than between component for the overall returns to English.

In the right panel of Appendix Table 2, we present the decompositions for women. Women's between-occupation share of the overall returns to English is greater than men's, which is consistent with more sorting into occupations on the basis of English skill for women.

6 Conclusion

In India, the raw difference in earnings between people who speak English and people who do not is large, but this overstates the economic value of English because higher ability people are more likely to be proficient in English. In this paper, we take advantage of a rich data set to mitigate this ability bias. After controlling for age, social group, schooling, geography and proxies for ability, we find that there are large, statistically significant returns to English-language skills in India. Hourly wages are on average 34% higher for men who speak fluent English and 13% higher for men who speak a little English relative to men who speak no English. For women, the average returns are 22% for fluent English and 10% for a little English. There is considerable heterogeneity in the returns to English. More experienced

---

[38]We added detailed industry dummies–there are about eighty industry codes–to the base model and observed less reduction in the estimated coefficients for English skill.

and more educated workers receive higher returns to English. The complementarity between English skills and education appears to have strengthened over time–only the more educated among young workers earn a premium for English skill, whereas older workers across all education groups do.

In India and many other developing countries, there is active debate over whether to promote the local language or a more globally accepted language like English in schools. While promoting the local language might make primary schooling more accessible and strengthen national identity, it may reduce economic opportunities because of the special role of English in the global economy. While we do find a positive association between English proficiency and earnings in India on average, we also find that returns are considerably more limited for more recent entrants into the labor market. Indeed, among workers who have not completed their secondary schooling, those who are more English proficient do not earn more. For the marginal worker deciding whether to invest in English skill, the relevant return would be the one estimated for the recent entrants. Thus, if we took our estimates as reflective of the causal returns to English–see Section 3 for the caveats–the implication is that providing English classes to adults may not necessarily raise their wages (as English skill raises wages only when coupled with high education). Even without the causal interpretation, our results highlight that individuals and policymakers should consider language-skill complementarities when making human capital investment decisions.

This study is the first to use a nationally representative sample to estimate the economic returns to English-language skills in India. Besides providing estimates of the average wage differences by English proficiency, we uncovered some patterns in the returns that were previously undocumented for India and worth further exploration, such as the language-skill complementarity, lower returns for scheduled castes, and lower returns in places with greater historical English prevalence. Another area for future work is the impact of English-language skills on social outcomes; the effects of English-language skills can extend beyond the labor market, and in ways that dramatically affect society.

References

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 113 (1): 151–184.

Angrist, Joshua, Aimee Chin, and Ricardo A. Godoy. 2008. "Is Spanish-Only Schooling Responsible for the Puerto Rican Language Gap?." *Journal of Development Economics* 85 (February): 105–108.

Angrist, Joshua and Victor Lavy. 1997. "The Effect of a Change in Language of Instruction on the Returns to Schooling in Morocco." *Journal of Labor Economics* 15 (January), no. 1: S48–76.

Berman, Eli, Kevin Lang, and Erez Siniver. 2003. "Language-Skill Complementarity: Returns to Immigrant Language Acquisition." *Labour Economics* 10 (June), no. 3: 265–290.

Bleakley, Hoyt and Aimee Chin. 2004. "Language Skills and Earnings: Evidence from Childhood Immigrants." *Review of Economics and Statistics* 86 (May), no. 2: 481–496.

Card, David. 1999. "The Causal Effect of Education on Earnings." In *Handbook of Labor Economics*, ed. Orley Ashenfelter and David Card. Amsterdam, North-Holland: Elsevier Science.

Clingingsmith, David. 2007. "Bilingualism, Language Shift, and Industrialization in Mid-20th Century India." Unpublished manuscript, Weatherhead School of Management, Case Western Reserve University.

Desai, Sonalde, Reeve Vanneman, and National Council of Applied Economic Research, New Delhi. 2005. India Human Development Survey (IHDS), [Computer file]. ICPSR22626-v5. Ann Arbor, MI: Inter-University Consortium for Political and Social Research [distributor], 2009-06-22.

Dustmann, Christian and Arthur van Soest. 2002. "Language and the Earnings of Immigrants." *Industrial and Labor Relations Review* 55, no. 3: 473–492.

He, Fang, Leigh L. Linden, and Margaret MacLeod. 2008. "How to Teach English in India: Testing the Relative Productivity of Instruction Methods within the Pratham English Language Education Program." Unpublished manuscript, Columbia University.

Hohenthal, Annika. 2003. "English in India: Loyalty and Attitudes." *Language in India* (May), no. 3.

Jensen, Robert T. 2010. "Economic Opportunities and Gender Differences in Human Capital: Experimental Evidence for India." NBER Working Paper no. 16021 (May). National Bureau of Economic Research, Cambridge, MA.

Kapur, Shilpi and Tanika Chakraborty. 2008. "English Language Premium: Evidence from a Policy Experiment in India." Unpublished manuscript, Department of Economics, Washington University in St. Louis.

Lang, Kevin and Erez Siniver. 2006. "The Return to English in a Non-English Speaking Country: Russian Immigrants and Native Israelis in Israel." NBER Working Paper 12464 (August). National Bureau of Economic Research, Cambridge, MA.

Levinsohn, James. 2007. "Globalization and the Returns to Speaking English in South Africa." In *Globalization and Poverty*, ed. Ann Harrison. Chicago: University of Chicago Press.

Munshi, Kaivan and Mark Rosenzweig. 2006. "Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy." *American Economic Review* 96 (September), no. 4: 1225–1252.

————. 2009. "Why is Mobility in India So Low? Social Insurance, Inequality and Growth." NBER Working Paper no. 14850 (April). National Bureau of Economic Research, Cambridge, MA.

Oster, Emily and M. Bryce Millett. 2010. "Do Call Centers Promote School Enrollment? Evidence from India." NBER Working Paper no. 15922 (April). National Bureau of Economic Research, Cambridge, MA.

Roy, Joydeep. 2004. "Can Changing Academic Standards Affect Educational Outcomes? Evidence from a Policy Experiment in India." Unpublished manuscript, Georgetown University.

Saiz, Albert and Zoido, Elena. 2005. "Listening to What the World Says: Bilingualism and Earnings in the United States." *Review of Economics and Statistics* 87 (January), no. 3: 523–538

Shastry, Gauri K. 2011. "Human Capital Response to Globalization: Education and Information Technology in India." Unpublished manuscript, Department of Economics, Wellesley College.

Figure 1. English-Speaking Ability Across the Indian States
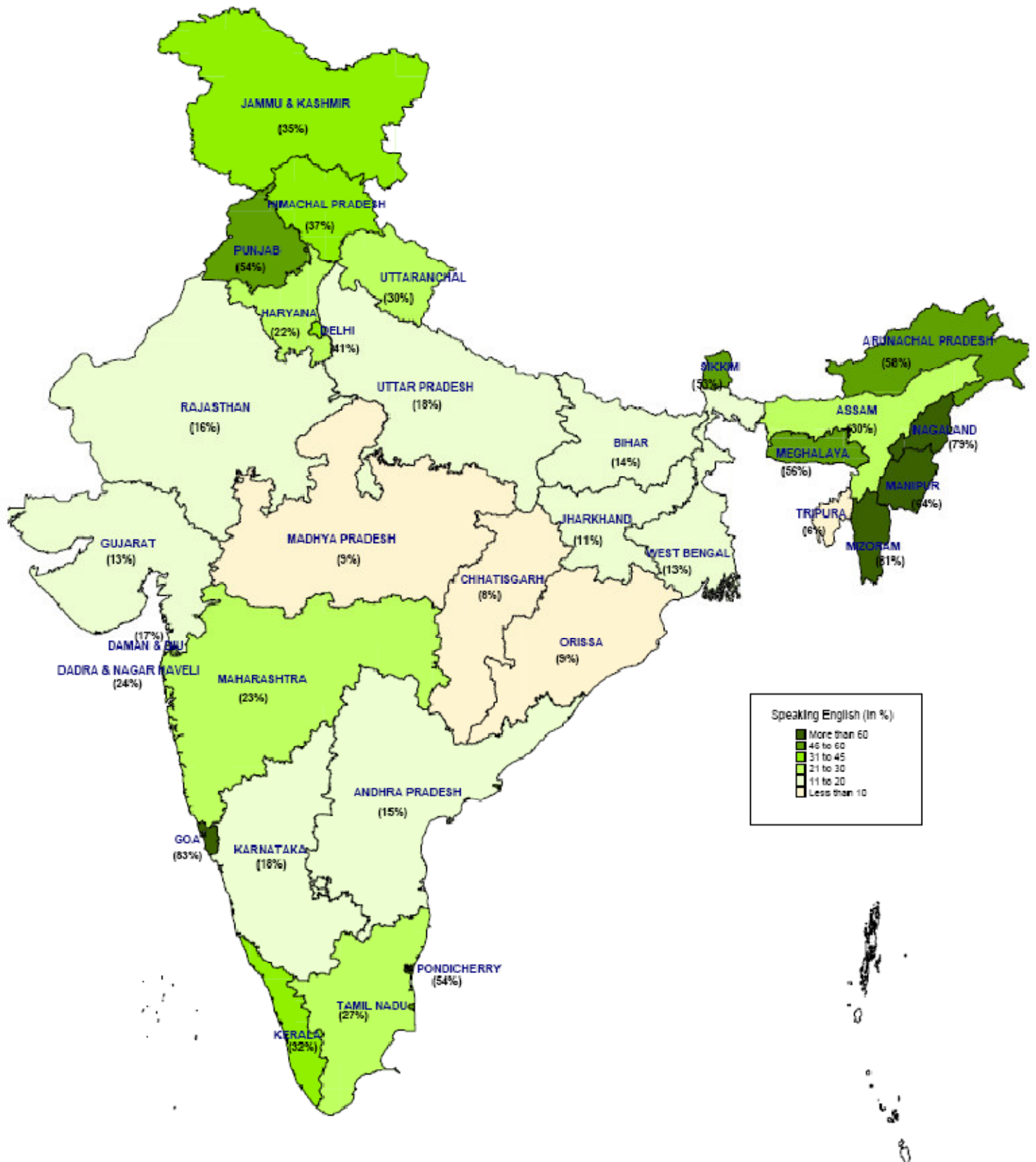
Notes: Reported in each state is the mean percentage speaking English (either fluently or a little). These means are calculated from data on individuals aged 18-65 from the India Human Development Survey, 2005.

Figure 2. Summary of Returns to English by Age and Education, Men
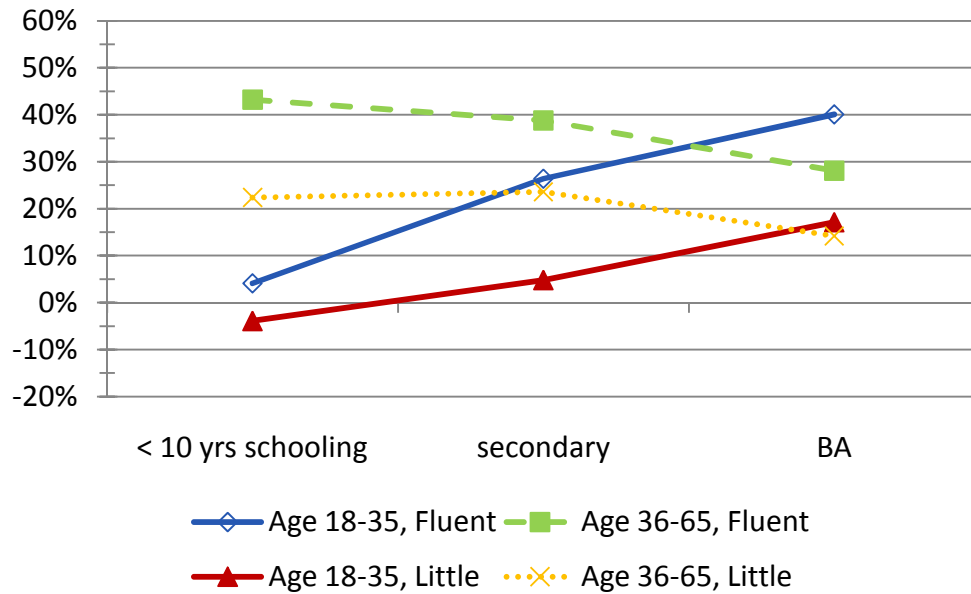
Notes: The returns are calculated based on the coefficients reported in Table 5, Column 4.

Table 1. Adult English-Speaking Ability in India

	Percent who converse fluently in English (1)	Percent who converse a little in English (2)	Total percent with English ability, (1)+(2) (3)
All Individuals Aged 18-65	3.8	16.2	20.0
	(19.1)	(36.9)	(40.0)
By Sex:			
Male	5.0	21.0	25.9
	(21.7)	(40.7)	(43.8)
Female	2.6	11.5	14.2
	(16.0)	(31.9)	(34.9)
By Age:			
18-35	4.4	20.2	24.6
	(20.4)	(40.2)	(43.0)
36-50	3.3	12.8	16.1
	(17.8)	(33.4)	(36.7)
51-65	3.0	9.7	12.7
	(16.9)	(29.6)	(33.2)
By Educational Attainment:			
No completed schooling	0.0	0.0	0.0
	(0.6)	(1.8)	(1.9)
Some primary (1-4 years schooling)	0.1	1.1	1.2
	(2.8)	(10.6)	(11.0)
Primary (5-9)	0.5	10.3	10.8
	(6.8)	(30.4)	(31.1)
Secondary (10-14)	6.0	49.5	55.5
	(23.7)	(50.0)	(49.7)
College graduate or higher	35.3	53.2	88.5
	(47.8)	(49.9)	(31.8)
By Social Group:			
Scheduled tribes	2.5	8.4	10.9
	(15.7)	(27.7)	(31.2)
Scheduled castes	1.5	11.4	12.9
	(12.3)	(31.7)	(33.5)
Other backward castes	2.9	14.7	17.6
	(16.9)	(35.4)	(38.1)
Higher castes	7.6	26.0	33.6
	(26.5)	(43.9)	(47.2)
By Geography:			
Urban	9.0	25.8	34.8
	(28.6)	(43.8)	(47.6)
Rural	1.7	12.4	14.1
	(12.9)	(32.9)	(34.8)
1961 English prevalence >= median	5.5	20.4	25.9
	(22.8)	(40.3)	(43.8)
1961 English prevalence < median	2.0	12.3	14.3
	(14.0)	(32.8)	(35.0)
1961 Linguistic heterogeneity >= median	4.2	17.2	21.5
	(20.2)	(37.8)	(41.1)
1961 Linguistic heterogeneity < median	3.3	15.4	18.7
	(17.8)	(36.1)	(39.0)
2003 IT presence: Yes	7.7	22.3	30.0
	(26.6)	(41.6)	(45.8)
2003 IT presence: No	2.6	14.3	16.9
	(15.8)	(35.0)	(37.5)

Notes: The sample consists of individuals aged 18-65 from the India Human Development Survey, 2005 (121,988 observations in top row, fewer in other rows because they are subsamples of the top row's sample). Standard deviations are shown in parentheses below the mean percent of given subpopulation with English ability. 1961 English prevalence is measured as the share of the individual's district population reporting English as a mother tongue or a second language in the 1961 Census. 1961 linguistic heterogeneity is measured using one minus the Herfindahl index of mother tongues in the individual's district, i.e., $1-\Sigma x_l^2$ where l indexes language and x is the share of the individual's district population reporting l as the mother tongue in the 1961 Census. IT presence is measured as a dummy variable indicating whether the individual's district had any IT firm headquarters or branch according to the 2003 National Association of Software and Service Companies (NASSCOM) directory.

Table 2. Summary Statistics

	All men in wage employment (1)	Male household heads in wage employment (2)	All male household heads (3)	All women in wage employment (4)
Log hourly wage last year	2.250	2.308		1.620
	(0.763)	(0.800)		(0.679)
Log household income last year			10.322	
			(0.980)	
Log monthly per capita consumption last year			6.497	
			(0.678)	
Fluent English	0.046	0.045	0.040	0.026
	(0.210)	(0.208)	(0.197)	(0.159)
Little English	0.165	0.150	0.167	0.060
	(0.372)	(0.357)	(0.373)	(0.238)
Urban residence	0.292	0.305	0.291	0.178
	(0.455)	(0.461)	(0.454)	(0.382)
Age	36.748	41.908	44.157	36.366
	(11.730)	(10.198)	(10.998)	(11.286)
Social group: scheduled tribes	0.095	0.100	0.083	0.152
	(0.293)	(0.300)	(0.276)	(0.359)
Social group: scheduled castes	0.271	0.275	0.223	0.313
	(0.445)	(0.446)	(0.416)	(0.464)
Social group: other backward castes	0.338	0.334	0.354	0.355
	(0.473)	(0.472)	(0.478)	(0.479)
Social group: Muslims	0.110	0.109	0.115	0.058
	(0.313)	(0.311)	(0.319)	(0.233)
Years of schooling completed: 0	0.297	0.351	0.316	0.677
	(0.457)	(0.477)	(0.465)	(0.468)
Years of schooling completed: 1	0.006	0.007	0.006	0.004
	(0.079)	(0.082)	(0.075)	(0.061)
Years of schooling completed: 2	0.023	0.026	0.026	0.018
	(0.151)	(0.159)	(0.160)	(0.133)
Years of schooling completed: 3	0.030	0.035	0.034	0.021
	(0.170)	(0.184)	(0.181)	(0.143)
Years of schooling completed: 4	0.044	0.047	0.049	0.027
	(0.206)	(0.211)	(0.216)	(0.163)
Years of schooling completed: 5 (Primary school completion)	0.085	0.089	0.089	0.055
	(0.279)	(0.284)	(0.285)	(0.227)
Years of schooling completed: 6	0.032	0.030	0.031	0.016
	(0.177)	(0.170)	(0.173)	(0.126)
Years of schooling completed: 7	0.055	0.049	0.051	0.032
	(0.227)	(0.216)	(0.220)	(0.177)
Years of schooling completed: 8 (Middle school completion)	0.069	0.061	0.067	0.023
	(0.253)	(0.239)	(0.250)	(0.149)
Years of schooling completed: 9	0.086	0.063	0.071	0.023
	(0.280)	(0.244)	(0.257)	(0.151)
Years of schooling completed: 10 (Secondary school completion)	0.106	0.099	0.109	0.030
	(0.308)	(0.298)	(0.312)	(0.170)
Years of schooling completed: 11	0.017	0.014	0.017	0.005
	(0.130)	(0.118)	(0.128)	(0.071)
Years of schooling completed: 12 (Higher secondary school completion)	0.058	0.048	0.054	0.023
	(0.234)	(0.215)	(0.225)	(0.150)
Years of schooling completed: 13	0.002	0.002	0.002	0.002
	(0.046)	(0.040)	(0.047)	(0.040)
Years of schooling completed: 14	0.003	0.002	0.003	0.002
	(0.056)	(0.047)	(0.051)	(0.041)
Years of schooling completed: 15 (Bachelor's degree or more)	0.086	0.077	0.075	0.043
	(0.280)	(0.267)	(0.263)	(0.202)
Performance on SSLC exam: Class I	0.056	0.050	0.048	0.033
	(0.231)	(0.219)	(0.214)	(0.178)
Performance on SSLC exam: Class II	0.141	0.125	0.139	0.048
	(0.348)	(0.331)	(0.346)	(0.214)
Performance on SSLC exam: Class III	0.049	0.041	0.046	0.014
	(0.215)	(0.199)	(0.210)	(0.119)
Has repeated or failed a grade	0.170	0.142	0.143	0.063
	(0.376)	(0.349)	(0.350)	(0.243)
Number of observations	32,445	21,057	33,376	12,205

Notes: The sample consists of individuals aged 18-65 from the India Human Development Survey, 2005 with non-missing English ability and educational attainment variables. In Column 3, due to missing values, there are fewer observations for two variables, household income (32,916) and household consumption (33,340). Standard deviations are shown in parentheses below the means.

Table 3. Effect of English-Speaking Ability on Male Log Hourly Wages

	No Controls (1)	Demographic Controls (2)	Education Controls (3)	Geographic Controls (4)	'Ability' Controls (5)	District Fixed Effects (6)
Fluent English	1.370***	1.201***	0.603***	0.393***	0.339***	0.345***
	(0.0230)	(0.0228)	(0.0301)	(0.0302)	(0.0317)	(0.0301)
Little English	0.690***	0.607***	0.224***	0.137***	0.125***	0.129***
	(0.0168)	(0.0160)	(0.0186)	(0.0201)	(0.0214)	(0.0186)
Age		0.041***	0.040***	0.032***	0.032***	0.031***
		(0.00242)	(0.00235)	(0.00221)	(0.00222)	(0.00212)
Age squared/100		-0.039***	-0.035***	-0.027***	-0.027***	-0.025***
		(0.00313)	(0.00304)	(0.00289)	(0.00290)	(0.00275)
Social group: scheduled castes		-0.335***	-0.210***	-0.125***	-0.118***	-0.102***
		(0.0146)	(0.0143)	(0.0133)	(0.0131)	(0.0130)
Social group: other backward castes		-0.303***	-0.228***	-0.119***	-0.116***	-0.085***
		(0.0146)	(0.0142)	(0.0136)	(0.0137)	(0.0132)
Social group: scheduled tribes		-0.529***	-0.385***	-0.237***	-0.231***	-0.167***
		(0.0172)	(0.0173)	(0.0180)	(0.0177)	(0.0175)
Social group: Muslims		-0.236***	-0.102***	-0.158***	-0.156***	-0.112***
		(0.0172)	(0.0166)	(0.0163)	(0.0162)	(0.0164)
Years of schooling completed: 1			0.069	0.0606	0.066*	0.069*
			(0.0425)	(0.0396)	(0.0397)	(0.0410)
Years of schooling completed: 2			0.041	0.037	0.041	0.056**
			(0.0268)	(0.0251)	(0.0252)	(0.0230)
Years of schooling completed: 3			0.047*	0.017	0.027	0.044*
			(0.0267)	(0.0268)	(0.0263)	(0.0231)
Years of schooling completed: 4			0.092***	0.048***	0.057***	0.075***
			(0.0188)	(0.0174)	(0.0176)	(0.0166)
Years of schooling completed: 5 (Primary school completion)			0.187***	0.123***	0.133***	0.122***
			(0.0161)	(0.0143)	(0.0142)	(0.0145)
Years of schooling completed: 6			0.228***	0.128***	0.140***	0.148***
			(0.0242)	(0.0218)	(0.0218)	(0.0210)
Years of schooling completed: 7			0.266***	0.185***	0.200***	0.191***
			(0.0209)	(0.0201)	(0.0202)	(0.0193)
Years of schooling completed: 8 (Middle school completion)			0.336***	0.217***	0.234***	0.232***
			(0.0188)	(0.0180)	(0.0185)	(0.0175)
Years of schooling completed: 9			0.352***	0.239***	0.266***	0.254***
			(0.0181)	(0.0173)	(0.0179)	(0.0167)
Years of schooling completed: 10 (Secondary school completion)			0.520***	0.377***	0.373***	0.334***
			(0.0194)	(0.0206)	(0.0605)	(0.0408)
Years of schooling completed: 11			0.360***	0.334***	0.325***	0.292***
			(0.0387)	(0.0364)	(0.0638)	(0.0513)
Years of schooling completed: 12 (Higher secondary school completion)			0.579***	0.467***	0.435***	0.410***
			(0.0265)	(0.0264)	(0.0593)	(0.0453)
Years of schooling completed: 13			0.621***	0.529***	0.496***	0.518***
			(0.105)	(0.0997)	(0.107)	(0.106)
Years of schooling completed: 14			0.605***	0.536***	0.490***	0.474***
			(0.106)	(0.109)	(0.122)	(0.110)
Years of schooling completed: 15 (Bachelor's degree or more)			0.913***	0.796***	0.720***	0.679***
			(0.0286)	(0.0283)	(0.0607)	(0.0457)
Performance on SSLC exam: Class I					0.227***	0.236***
					(0.0687)	(0.0494)
Performance on SSLC exam: Class II					0.065	0.070
					(0.0640)	(0.0456)
Performance on SSLC exam: Class III					-0.094	-0.063
					(0.0646)	(0.0465)
Has repeated or failed a grade					-0.061***	-0.069***
					(0.0155)	(0.0128)
Urban residence				0.413***	0.405***	0.341***
				(0.0110)	(0.0108)	(0.0134)
State fixed effects	No	No	No	Yes	Yes	No
District fixed effects	No	No	No	No	No	Yes
R-squared	0.231	0.299	0.360	0.455	0.459	0.516

Notes: The sample consists of men aged 18-65 who reported wage and salary work from the India Human Development Survey, 2005 (32,445 observations in all columns). The omitted social group consists of the high castes (88% of this omitted group), 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***	0.312***	0.337***	0.320***	0.257***	0.235***
	(0.0378)	(0.0380)	(0.0365)	(0.0364)	(0.0231)	(0.0232)
Little English	0.163***	0.160***	0.113***	0.106***	0.103***	0.098***
	(0.0238)	(0.0237)	(0.0240)	(0.0239)	(0.0140)	(0.0142)
R-squared	0.556	0.557	0.369	0.373	0.468	0.473
Number of observations	21,057	20,672	32,916	32,272	33,340	32,688

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 (there were a few missing values for this variable). Robust standard errors are shown in parentheses.
Asterisks denote significance levels (*=.10, **=.05, ***=.01).

Table 5. Heterogeneity in Effect of English-Speaking Ability on Male Wages by Demographic Variables

	Base	Age	Education	Age-Education	Age-Education	Social Group	Soc Gp + Age-Educ
Variable 1:		36-50	BA	Age 36-65	Age 36-65	ST	ST
Variable 2:		51-65	Secondary	BA	BA	SC	SC
Variable 3:				Secondary		OBC	OBC
Omitted group:		*Age 18-35*	*< 10 yrs schooling*	*Lower ed & younger*	*Lower ed & younger*	*Higher castes*	*Higher castes*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fluent English	0.345***	0.113***	0.241**	0.041	0.134**	0.395***	0.180***
	(0.0301)	(0.0369)	(0.108)	(0.172)	(0.0523)	(0.0345)	(0.0567)
Little English	0.129***	-0.074***	0.069***	-0.039	-0.060***	0.150***	-0.056
	(0.0186)	(0.0180)	(0.0245)	(0.0249)	(0.0188)	(0.0232)	(0.0251)
Fluent English x Variable 1		0.376***	0.123	0.391*	0.401***	-0.042	-0.018
		(0.0439)	(0.121)	(0.217)	(0.0719)	(0.113)	(0.115)
Little English x Variable 1		0.369***	0.095	0.263***	0.406***	0.057	0.084
		(0.0267)	(0.0606)	(0.0453)	(0.0301)	(0.0643)	(0.0633)
Fluent English x Variable 2		0.565***	0.155	0.360*	0.267***	-0.180**	-0.141**
		(0.0554)	(0.115)	(0.186)	(0.0881)	(0.0708)	(0.0656)
Little English x Variable 2		0.611***	0.091***	0.210***	0.231***	-0.094***	-0.042
		(0.0486)	(0.0352)	(0.0733)	(0.0714)	(0.0322)	(0.0311)
Fluent English x Variable 3				0.223		-0.075	-0.065
				(0.179)		(0.0500)	(0.0511)
Little English x Variable 3				0.087**		-0.011	0.006
				(0.0357)		(0.0351)	(0.0331)
Fluent English x BA x Age 36-65				-0.511**	-0.523***		Not reported
				(0.241)	(0.126)		
Little English x BA x Age 36-65				-0.292**	-0.436***		Not reported
				(0.114)	(0.109)		
Fluent English x Secondary School x Age 36-65				-0.267			
				(0.231)			
Little English x Secondary School x Age 36-65				-0.0749			
				(0.0642)			

Notes: The sample consists of men aged 18-65 who reported wage and salary work from the India Human Development Survey, 2005 (32,445 observations in all columns). 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). Columns 2, 4, 5 and 7 additionally control for a dummy for being age 36-50 and one for being 51-65. In Column 7, we start with the Column 6 specification and control for age-education-specific effects of English skill as in Column 5 (we have suppressed reporting the coefficients for the age and education interactions with English skill, so we should note that the main effects of English skill reflect the returns for lower educated, younger and higher caste men). Robust standard errors are shown in parentheses. Asterisks denote significance levels (*=.10, **=.05, ***=.01).

Table 6. Heterogeneity in Effect of English-Speaking Ability on Male Wages by Geographic Variables

	Base (1)	Urban/ Rural (2)	1961 English Prevalence (3)	1961 Linguistic Heterogeneity (4)	2003 IT Presence (5)	All Four Geog. Variables (6)	(6) + Age-Educ Interactions (7)
Fluent English	0.345*** (0.0301)	0.361*** (0.0523)	0.374*** (0.0323)	0.358*** (0.0311)	0.413*** (0.0395)	0.387*** (0.0543)	0.170** (0.0682)
Little English	0.129*** (0.0186)	0.135*** (0.0256)	0.136*** (0.0191)	0.129*** (0.0189)	0.130*** (0.0228)	0.125*** (0.0271)	-0.0505** (0.0231)
Fluent English x Urban		-0.028 (0.0537)				0.0433 (0.0561)	0.0292 (0.0563)
Little English x Urban		-0.015 (0.0286)				-0.00298 (0.0309)	-0.0574* (0.0307)
Fluent English x District Share English			-1.783*** (0.396)			-1.340*** (0.435)	-1.456*** (0.440)
Little English x District Share English			-1.222*** (0.280)			-1.379*** (0.318)	-1.435*** (0.301)
Fluent English x District Linguistic Heterogeneity				-0.254** (0.103)		-0.119 (0.104)	-0.109 (0.104)
Little English x District Linguistic Heterogeneity				-0.0544 (0.0692)		-0.0118 (0.0725)	-0.00951 (0.0690)
Fluent English x District has IT					-0.147*** (0.0433)	-0.0981** (0.0472)	-0.0685 (0.0466)
Little English x District has IT					-0.00541 (0.0303)	0.0390 (0.0344)	0.0546* (0.0324)

Notes: The sample consists of men aged 18-65 who reported wage and salary work from the India Human Development Survey, 2005 (32,445 observations in all columns except Columns 3, 4, 6 and 7, which have 31,919 observations due to lack of 1961 Census language information for a few districts). 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). In Column 7, we start with the Column 6 specification and control for age-education-specific effects of English-skill and dummies for being age 36-50 and one for being 51-6 as in Table 5, Column 5. English prevalence is measured as the share (from 0 to 1) of the individual's district population reporting English as a mother tongue or a second language in the 1961 Census. Linguistic heterogeneity is measured using one minus the Herfindahl index of mother tongues in the individual's district, i.e., $1-\Sigma x_l^2$ where l indexes language and x is the share of the individual's district population reporting l as the mother tongue in the 1961 Census (varies from 0 to 1, with 1 being the most heterogeneous). IT presence is measured as a dummy variable indicating whether the individual's district had any IT firm headquarters or branch according to the 2003 National Association of Software and Service Companies (NASSCOM) directory. Robust standard errors are shown in parentheses. Asterisks denote significance levels (*=.10, **=.05, ***=.01).

Table 7. Effect of English-Speaking Ability on Female Wages

	Base	**Age-Education**	**Social Group + Age-Educ Interactions**	**Geography + Age-Educ Interactions**
Variable 1:		Age 36-65	ST	Urban
Variable 2:		BA	SC	English Prevalence
Variable 3:			OBC	Linguistic Heterogeneity
Variable 4:				IT Presence
	(1)	(2)	(3)	(4)
Fluent English	0.223***	-0.116	-0.0441	-0.214*
	(0.0775)	(0.106)	(0.117)	(0.127)
Little English	0.100**	-0.093*	0.0516	-0.0840
	(0.0511)	(0.0523)	(0.0712)	(0.0566)
Fluent English x Variable 1		0.602***	0.0255	0.253*
		(0.185)	(0.139)	(0.150)
Little English x Variable 1		0.533***	0.135	0.133*
		(0.0893)	(0.124)	(0.0734)
Fluent English x Variable 2		0.663***	-0.123	-2.116*
		(0.194)	(0.151)	(1.270)
Little English x Variable 2		0.381**	-0.235***	-3.951***
		(0.168)	(0.0797)	(0.907)
Fluent English x Variable 3			-0.101	0.212
			(0.126)	(0.231)
Little English x Variable 3			-0.228***	0.319**
			(0.0750)	(0.158)
Fluent English x Variable 4				0.0240
				(0.121)
Little English x Variable 4				-0.0155
				(0.0815)
Fluent English x BA x Age 36-65		-0.997***	Not reported	Not reported
		(0.332)		
Little English x BA x Age 36-65		-0.591**	Not reported	Not reported
		(0.289)		

Notes: The sample consists of women aged 18-65 who reported wage and salary work from the India Human Development Survey, 2005 (12,205 observations in all columns except Column 4, which has 12,093 observations due to lack of 1961 Census language information for a few districts). Each column reports the results of a separate regression. Column 1 uses the same specification as Table 3, Column 6; Column 2 as Table 5, Column 5; Column 3 as Table 5, Column 7; and Column 4 as Table 6, Column 7. Robust standard errors are shown in parentheses. Asterisks denote significance levels (*=.10, **=.05, ***=.01).

Appendix Table 1. Effect of English-Speaking Ability on the Probability of Employment

	Male is in wage employment (1)	Male HH head is in wage employment (2)	Male HH head is in any type of employment (3)	Female aged 18-65 is in wage employment (4)
Fluent English	0.0761***	0.134***	-0.0006	0.170***
	(0.0193)	(0.0179)	(0.009)	(0.0257)
Little English	0.0073	0.0162	-0.0101	0.0435***
	(0.0102)	(0.0134)	(0.006)	(0.0156)
Sample	Males aged 18-65	Males aged 18-65 who are HH heads	Males aged 18-65 who are HH heads	Females aged 18-65
Number of observations	61,062	33,371	32,642	60,549
Mean (s.d.) of dependent variable	0.540 (0.498)	0.631 (0.483)	0.908 (0.289)	0.216 (0.411)

Notes: Data are from the India Human Development Survey, 2005. Each column reports the marginal effects from a separate probit 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). Though our sample has 33,376 male household heads and 60,816 women, the reported number of observations Columns 2-4 is less because a few districts had no variation in the dependent variable. Robust standard errors are shown in parentheses. Asterisks denote significance levels (*=.10, **=.05, ***=.01).

Appendix Table 2. Role of Occupation and Industry Choice

	Men				Women			
	Base (1)	(1) + Occupation Dummies (2)	(1) + Industry Dummies (3)	(1) + Occup & Ind Dummies (4)	Base (5)	(5) + Occupation Dummies (6)	(5) + Industry Dummies (7)	(5) + Occup & Ind Dummies (8)
Fluent English	0.345 ***	0.223 ***	0.281 ***	0.216 ***	0.223 ***	0.065	0.157 **	0.096
	(0.0301)	(0.0282)	(0.0288)	(0.0282)	(0.0775)	(0.0740)	(0.0770)	(0.0761)
% of Base Effect		*65%*	*81%*	*63%*		*29%*	*70%*	*43%*
Little English	0.129 ***	0.072 ***	0.091 ***	0.068 ***	0.100 **	0.023	0.092 *	0.057
	(0.0186)	(0.0157)	(0.0157)	(0.0151)	(0.0511)	(0.0475)	(0.0507)	(0.0504)
% of Base Effect		*56%*	*71%*	*53%*		*23%*	*92%*	*57%*
R-squared	0.516	0.596	0.587	0.611	0.550	0.599	0.592	0.608
Number of observations	32,445	32,391	30,256	30,245	12,205	12,187	11,829	11,824

Notes: The sample consists of men aged 18-65 who reported wage and salary work from the India Human Development Survey, 2005.
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 3, Column 6). Robust standard errors are shown in parentheses.
Asterisks denote significance levels (*=.10, **=.05, ***=.01).