# Earning but Not Learning? English Medium Instruction in Indian Primary Schools

**Khushboo Gupta[a] and Panu Pelkonen[b,c,d]**

[a] IFMR LEAD

[b] Department of Economics, University of Sussex

[c] Centre for the Economics of Education (CEE), London School of Economics

[d] IZA

12th January, 2017

**Abstract**: English-medium instruction has become increasingly popular in India due to high labour market returns of English language skills. We estimate the effect that the English-medium instruction has on learning outcomes in reading and mathematics, using a cross-section of pupil-level test scores. We confirm that higher socio-economic groups are more likely to opt for English medium schools. However, after controlling for a rich set of household characteristics and district effects, we show that the annual value added is much lower in English medium schools, in both public and private sector schools, for both reading and mathematics. We also show that the low value added in English medium schools is partly mediated in districts where private school teachers have higher reported qualifications in English.

Keywords: Medium of instruction, Language, Schooling, India. JEL Codes: I21, J18, J24

* Correspondence: Panu Pelkonen, Department of Economics, University of Sussex, Brighton, BN1 9RH, UK. p.o.pelkonen@sussex.ac.uk

Running title: English medium instruction in India

## 1. Introduction

English language has always held a special status in education system of India due to its usage in various levels of government and mass media. Studies show a high wage return in India for fluency in English (Azam et al, 2013, Chakraborty and Bakshi, 2016). In response to the increasing demand for English language, many English medium private schools have been established all over the country. In some states, some government schools also offer English as the medium of instruction. The Sixth all India Education survey conducted in 1993 found that English was used as a medium of instruction in 4.99% schools at primary stage, and by 2002, the corresponding figures had risen to 12.98%.

There are not many studies that look at the relationship between language of instruction and student learning; and to our knowledge, no generalizable studies have been conducted in India.[1] For other countries, the effect of the medium of instruction on learning has been studied, with mixed results (Bühmann and Trudell (2008) for Mali, Gfeller and Robinson (1998) for Cameroon, Angrist and Lavy (1997) for Morocco and Yip et al (2003) for Hong Kong).

In this study, we show that in India, there is a positive selection of pupils into schools that teach in English, in both public and private sector. However, after controlling for a rich set of socioeconomic characteristics and local area effects, we find that yearly value added in both reading and mathematics is substantially lower in English medium schools. The result holds for both public and private schools.

---

[1] The only work similar to ours we are aware of is a recent working paper by Nair (2015), who estimates the effect of medium of instruction on learning in Andhra Pradesh. His data has a longitudinal dimension, and his results broadly support those of ours.

A register database of Indian primary schools allows us to measure the level of English qualifications for private sector primary school teachers in each Indian district. Using this data, we show that the disadvantage from English medium instruction at the primary level is greater in areas with lower level English language qualifications of teachers.

Section 2 describes the data, followed by the estimation in Section 3. Results are summarised in Section 4 and Section 5 provides the conclusions.

## 2. Data

The main source of data is the ASER survey conducted by ASER centre Delhi, India. This is a household-based survey which is conducted annually since 2005 in all rural districts in India. The sample size is 600 households per rural district – 30 villages per district and 20 households per village. This gives a sample of over 300,000 households and over half a million individual observations for the whole country. Children in the age group of 5-16 who reside in the sampled household are tested in basic reading and basic arithmetic.

For our analysis, we use the 2011 round of the survey since it includes data on both mother tongue and language of instruction at school. The sample size for the 2011 survey is a little over 550,000 children in the age-group of 5-16 in 560 districts in India. We drop observations for which data on any of the child characteristics, language variables, household or parental characteristics, mathematics or reading test scores is missing. We also delete those observations whose language of instruction is an Indian language different from their mother tongue. Previous studies have shown that mother-tongue education affects learning mostly during the first few years of schooling. We therefore restrict our analysis to the children enrolled in (lower) primary school from grade 1 to 5 or grade 1 to 4 depending on the state

education policy. We only include children aged 5 to 12 and our final sample has about 98,000 observations.

The outcome variables we examine are the standardized reading and math scores of the children. For the reading test, the child is categorized into the highest category that they can reach: those who can't read the alphabet, those who can read the alphabet, those who can read the word, those who can read the paragraph, and those who can read the short story. These categories are then coded 0 to 4 respectively. Similarly for the arithmetic test, the child is categorized into the highest category that they can reach: those who can't recognize numbers 1-9, those who can recognize numbers 1-9, those who can recognize numbers 11-99, those who can do 2 digit subtractions, those who do 3 digits by 1 digit division. These categories are then coded 0 to 4 respectively. To facilitate the interpretation of the results, we normalize both test scores by age.

Our primary group of interest is the group of children taught in English (18% of the full sample). We control for additional variables that include child characteristics such as school grade, type of school attended; parental years of schooling and household characteristics that measure household wealth. We also control for the child's age to grade with a set of dummies. Table 1 presents the summary statistics of these variables for the two language groups.

TABLE 1

The unconditional normalised skill difference between the groups across all ages is roughly half a standard deviation. About 24% of English medium students take extra-curricular paid tuition compared to 17% of students taught in mother tongue. English medium

students are more than twice as likely to be enrolled in a private school as students enrolled in mother tongue instruction. 73% of children in English medium choose English as the test language.[2] English medium students have, on average, more educated parents and higher ownership of household assets.

## 3. Estimation

The main challenge of identification is the selection of pupils into different schools. This applies to both the choice of private vs government school, but also the medium of instruction. No exogenous source of variation is available to us to explore the matter. Instead, we use the cross-sectional data to estimate the effect of English medium schools throughout the grades of primary schooling and argue that learning differences across different types of schools in the first grade are reflective of selection, while the differences in the gained skills across later grade levels are reflective of the value added that the different schools are creating. More precisely,

$$Y_{ij} = \alpha_0 + \alpha_1 Eng_{ij} + \alpha_2 Grade_{ij} + \alpha_3 Eng_{ij} \times Grade_{ij} + \theta X_{ij} + \lambda_j + \omega_{ij} , \qquad (1)$$

in which $Y_{ij}$ be the learning outcome as measured by the normalized test score of individual $i$ in district $j$. $Eng_{ij}$ is a dummy variable that takes the value 1 if the individual $i$ in district $j$ studies in an English language of instruction and 0 if the individual studies in his/her mother tongue. $X_{ij}$ is a vector of all other covariates including household characteristics, child

[2] When tested, the children in the household can choose the language of testing from the most common languages of the state, or English.

characteristics, school type, and parental years of schooling that could also affect the learning outcomes. We include district fixed effects ($\lambda_j$) so that all comparisons in learning levels between the two language groups occur within areas with the same education policies.[3]

In equation (1), the marginal effect of English medium is

$$\partial Y_{ij}/\partial Eng_{ij} = \alpha_1 + \alpha_3 Grade_{ij}. \quad (2)$$

Here, $\alpha_1 + \alpha_3$ is the estimate of the effect of English medium on skills in Grade 1. Since the pupils have at the time of testing been in school for less than half a year, we interpret this measure to largely capture the initial level of skills of the pupils, which should be overwhelmingly due to selection.

The average marginal effect of additional grade on learning, or annual value added, can be written as

$$\partial Y_{ij}/\partial Grade_{ij} = \alpha_2 + \alpha_3 Eng_{ij}. \quad (3)$$

Here, additional year of schooling increases skills by $\alpha_2$ for pupils in native medium schools and by $\alpha_2 + \alpha_3$ for pupils in English medium schools. Figure 1 provides a graphical version of the relationship between grades and learning among selected English medium pupils and other pupils.

[3] The education policies vary mostly at state level, but district level controls will account for local variations as well. The results are very similar with just state level fixed effects.

## 4. Results

Table 2 shows the results of equation 2, for both reading and mathematics, separately for government and private schools. The main estimates of interest are the English medium instruction ($\alpha_1$), grade level ($\alpha_2$), and the interaction between English medium and the grade level ($\alpha_3$). The controls include all variables listed in Table 1, as well as district effects.

TABLE 2

The results of Table 2 are interpreted using the marginal effects as discussed with equations (1) - (3), and the results are summarised in Table 3. We suggest that the selection of pupils in different schools could be plausibly well approximated by the differences in skills in Grade 1 ($\alpha_1 + \alpha_3$). Positive $\alpha_1 + \alpha_3$ suggests that pupils in English medium schools are positively selected. As Table 3 details, this is the case for both public and private schools, and for both reading and maths. In government sector the difference is .0517 standard deviations in reading and .0907 standard deviations in maths. In Private sector, the selection is stronger, with corresponding figures of .2285 for reading and .2782 for maths.

TABLE 3

Annual value added estimates are based on $\alpha_2$ for native medium pupils and $\alpha_2 + \alpha_3$ for English medium pupils. As Table 3 shows, the annual value added in government schools more than twice larger if the pupil is taught in native language, as opposed to English (.1002

vs .0429 in reading and .0972 vs .0389 in maths). In private sector schools, the difference is even more pronounced: The value added per year in English medium schools is tiny (under 0.03 standard deviations for both outcomes), but still statistically significantly different from zero. With native medium however, the value added is more than 0.10 standard deviations for both outcomes.

The results are robust to a non-parametric specification where the value added of each grade is estimated separately (results available on request). Ceiling effect of the test is very unlikely to explain the effects since only about half of 4$^{th}$ grade pupils have reached the highest reading level, and under a third has reached the highest mathematics level irrespective of medium of instruction. The main uncertainty over the results is due to the cross-sectional nature of the data.

### 4.1 Teacher qualifications in English in the private sector schools

Poor learning outcomes in English medium schools could be partly explained by poor knowledge of English by teachers in private primary schools. We must keep in mind that ASER is focused on rural areas, where use and visibility of English is lower than in cities.

Data on teachers' qualifications and skills are hard to come by, but one potential national-level source is the DISE database, which includes recognized private schools. Table 4 shows summary statistics of private school teachers' qualifications for 2011, calculated from DISE 2011 by district, and matched to ASER private schools by district. We see that the proportion of teachers who have studied English to university level is on average 25%. The average tenure of teachers across districts is 6.7 years, while 60% are typically graduates.

These data are self-reported by private schools and not verified independently. As such, they may contain measurement error.

TABLE 4

To assess whether teacher qualifications in English could be behind the poor value added of English medium schools, we split the districts in India to those that have better than median, or worse than median English qualifications for rural private school teachers, and re-estimate column 2 of Table 2 for both groups. The results are in Table 5. Comparing columns 1 and 2 shows that the value added measure $\alpha_2 + \alpha_3$ is virtually zero in districts with lower than median English qualifications, but it is positive for districts with better English qualifications.

TABLE 5

Columns 3 and 4 of the same table show that this result can't be replicated by comparing districts with shorter or longer tenure for teachers (assuming younger teachers are better in English). Columns 5 and 6 on the other hand, make a similar split to districts with higher and lower share of teacher who have a university degree. Again, this measure does not appear to be important in determining the value added for private school pupils who study in English.

**5. Concluding remarks**

A plausible reason for the negative effect of English medium on learning is the quality of teachers, especially with regards to their fluency in English. A study by NCERT (2012) on English language instruction in government schools in India found that most of the teachers were not qualified to teach English, had poor knowledge of grammar and inadequate writing and communication skills. For private schools, no such reports exist yet. Our results confirm that the yearly value added in English medium schools in near to zero in districts where private school teachers have poorer than average qualifications in English.

This topic is of particular relevance to India, where the percentage of schools using English language for instruction has been rising. However, until about two decades ago, English was used mainly by the educated middle class in the urban areas. Over the past two decades, there has been a rising demand for English skills in rural areas and among the economically weaker class. This has been partly because English is seen as an essential tool for economic prosperity. Despite poorer levels of learning in English medium schools, it may still be economical to study in English medium if the returns to English language skills are high.

**References**

Angrist, Joshua D., and Lavy, Victor. 1997. 'The Effect Of A Change In Language Of Instruction On The Returns To Schooling In Morocco'. *Journal Of Labor Economics* ,15 (s1): S48-S76.

Azam, Mehtabul, Chin, Aimee and Prakash, Nishith. 2013. 'The Returns To English-Language Skills In India'. *Economic Development And Cultural Change*, 61 (2): 335-367.

Bühmann, Dörthe, and Trudell, Barbara. 2008. 'Mother Tongue Matters: Local Language As A Key To Effective Learning.' Paris: UNESCO.

Chakrabort, Tanika, and Shilpi Kapur Bakshi. 2016. English Language Premium: Evidence from a policy Experiment in India. *Economics of Education Review*, forthcoming.

Gfeller, Elisabeth, and Robinson, Clinton. 1998. 'Which Language For Teaching? The Cultural Messages Transmitted By The Languages Used In Education'. *Language And Education* 12 (1): 18-32.

Nair, P. Sree. "Does Medium Of Instruction Affect Learning Outcomes?—Evidence Using Young Lives Longitudinal Data Of Andhra Pradesh, India". *PERI ESP Working Paper*, 1-17.

NCERT, 2006. Seventh All India Education Survey: Media Of Instruction And Languages Taught. New Delhi: NCERT.

NCERT, 2012. *Teaching Of English At Primary Level In Government Schools*. New Delhi: NCERT.

Yip, Din Yan, Wing Kwong Tsang, and Sin Pui Cheung. 2003. 'Evaluation Of The Effects Of Medium Of Instruction On The Science Learning Of Hong Kong Secondary Students: Performance On The Science Achievement Test'. *Bilingual Research Journal*, 27 (2): 295-331.

**Table 1 Summary statistics by language of instruction**

| | Native medium | | | | English medium | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | S.D | Min | Max | Mean | S.D | Min | Max |
| **Learning outcomes** | | | | | | | | |
| Raw reading score | 1.97 | 1.37 | 0 | 4 | 2.52 | 1.17 | 0 | 4 |
| Raw mathematics score | 1.75 | 1.18 | 0 | 4 | 2.41 | .98 | 0 | 4 |
| Normalised reading score | -.12 | 1.05 | -3.22 | 3.56 | .31 | .94 | -3.22 | 3.56 |
| Normalised mathematics score | -.16 | 1.01 | -2.89 | 3.79 | .43 | .87 | -2.89 | 3.79 |
| **Child and school** | | | | | | | | |
| Child age | 8.21 | 1.84 | 5 | 12 | 8.37 | 1.84 | 5 | 12 |
| Grade | 2.84 | 1.38 | 1 | 5 | 2.76 | 1.37 | 1 | 5 |
| Male | .539 | .498 | 0 | 1 | .542 | .498 | 0 | 1 |
| Pupils pays for tuition | .169 | .375 | 0 | 1 | .242 | .428 | 0 | 1 |
| School: Government school | .739 | .439 | 0 | 1 | .371 | .483 | 0 | 1 |
| School: Private school | .260 | .438 | 0 | 1 | .627 | .484 | 0 | 1 |
| Chose same test language as in school | .993 | .086 | 0 | 1 | .729 | .444 | 0 | 1 |
| **Home environment** | | | | | | | | |
| Father's years of education | 6.28 | 4.77 | 0 | 18 | 8.17 | 4.62 | 0 | 18 |
| Mother's years of education | 3.64 | 4.33 | 0 | 18 | 6.48 | 4.76 | 0 | 18 |
| Electricity | .648 | .477 | 0 | 1 | .944 | .231 | 0 | 1 |
| Toilet | .330 | .470 | 0 | 1 | .751 | .433 | 0 | 1 |
| Television | .429 | .495 | 0 | 1 | .684 | .465 | 0 | 1 |
| Mobile phone | .739 | .439 | 0 | 1 | .839 | .368 | 0 | 1 |
| Newspapers | .075 | .263 | 0 | 1 | .209 | .407 | 0 | 1 |
| Reading materials | .185 | .388 | 0 | 1 | .490 | .500 | 0 | 1 |
| Knowledge of computers | .069 | .254 | 0 | 1 | .207 | .405 | 0 | 1 |
| House: Pukka (solid) | .307 | .461 | 0 | 1 | .350 | .477 | 0 | 1 |
| House: Semi-Pukka | .354 | .478 | 0 | 1 | .355 | .478 | 0 | 1 |
| Observations | 80404 (81.9%) | | | | 17777 (18.1%) | | | |

**Table 2 Learning outcomes over grades and English medium**

| | [1] | [2] | [3] | [4] |
|---|---|---|---|---|
| | Reading | | Maths | |
| | Government | Private | Government | Private |
| English medium (α1) | .109* | .321** | .149** | .359** |
| | [.0436] | [.0345] | [.044] | [.0316] |
| Grade level (α2) | .1** | .119** | .0976** | .101** |
| | [.00272] | [.00449] | [.00261] | [.0044] |
| English X Grade level (α3) | -.0572** | -.0924** | -.0587** | -.0808** |
| | [.00729] | [.00699] | [.00705] | [.00671] |
| District controls | Y | Y | Y | Y |
| Child and school controls | Y | Y | Y | Y |
| Home environment controls | Y | Y | Y | Y |
| Age to grade controls | Y | Y | Y | Y |
| Parental education controls | Y | Y | Y | Y |
| Observations | 66151 | 32030 | 66151 | 32030 |
| R-squared | .329 | .294 | .350 | .329 |

Notes : ** p<0.01, * p<0.05, +p<0.1. Robust standard errors.

**Table 3 Summary of results**

| Panel A: Reading | | | |
|---|---|---|---|
| Sector | Parameter | Estimate | Interpretation |
| Government | $\alpha_1 + \alpha_3$ | .0517 (p = .2069) | Initial English vs Native difference in grade 1 |
| | $\alpha_2$ | .1002 (p = .0000) | Value added per grade, Native medium |
| | $\alpha_2 + \alpha_3$ | .0429 (p = .0000) | Value added per grade, English medium |
| Private | $\alpha_1 + \alpha_3$ | .2285 (p = .0000) | Initial English vs Native difference in grade 1 |
| | $\alpha_2$ | .1186 (p = .0000) | Value added per grade, Native medium |
| | $\alpha_2 + \alpha_3$ | .0262 (p = .0000) | Value added per grade, English medium |
| Panel B: Maths | | | |
| Sector | Parameter | Estimate | Interpretation |
| Government | $\alpha_1 + \alpha_3$ | .0907 (p = .0285) | Initial English vs Native difference in grade 1 |
| | $\alpha_2$ | .0976 (p = .0000) | Value added per grade, Native medium |
| | $\alpha_2 + \alpha_3$ | .0389 (p = .0000) | Value added per grade, English medium |
| Private | $\alpha_1 + \alpha_3$ | .2782 (p = .0000) | Initial English vs Native difference in grade 1 |
| | $\alpha_2$ | .1014 (p = .0000) | Value added per grade, Native medium |
| | $\alpha_2 + \alpha_3$ | .0206 (p = .0001) | Value added per grade, English medium |

**Table 4 Teacher qualifications and tenure for private primary schools in rural areas, 2011**

| | Obs. | Mean | S.D. | Min | Max |
|---|---|---|---|---|---|
| % studied English to university level | 31293 | .251 | .223 | 0 | .914 |
| Average years of tenure | 31293 | 6.67 | 3.27 | 1 | 23.29 |
| % Graduates | 31293 | .601 | .192 | 0 | .952 |

Source: DISE 2011 by district

**Table 5 English medium value added on reading in private schools and local level teacher qualifications**

| | [1] Low English quals | [2] High English quals | [3] Low tenure | [4] High tenure | [5] Low % graduates | [6] High % graduates |
|---|---|---|---|---|---|---|
| English medium (α1) | .295* | .251** | .302** | .362** | .339** | .29** |
| | [.147] | [.0411] | [.0428] | [.0663] | [.0504] | [.0499] |
| Grade level (α2) | .128** | .0981** | .106** | .13** | .116** | .122** |
| | [.00547] | [.00816] | [.00677] | [.00611] | [.00719] | [.00587] |
| English X Grade level (α3) | -.123** | -.0667** | -.0928** | -.083** | -.0934** | -.0883** |
| | [.0164] | [.00995] | [.00971] | [.0104] | [.00996] | [.0105] |
| District controls | Y | Y | Y | Y | Y | Y |
| Child and school controls | Y | Y | Y | Y | Y | Y |
| Home environment controls | Y | Y | Y | Y | Y | Y |
| Age to grade controls | Y | Y | Y | Y | Y | Y |
| Parental education controls | Y | Y | Y | Y | Y | Y |
| Observations | 15818 | 15475 | 15620 | 15673 | 15460 | 15833 |
| R-squared | .294 | .262 | .273 | .318 | .299 | .295 |

Notes : ** p<0.01, * p<0.05, +p<0.1. Robust standard errors. In pairs of columns, districts are split into groups of above or below median teacher characteristics.

**Figure 1 Relationship between grades and learning outcomes for selected and non-selected groups.**