



# **The effects of mediums of instruction on educational- and labor market outcomes: Evidence from Malaysia**

Rasyad Parinduri and Kian Ong

Nottingham University Business School, University of Nottingham  
Malaysia Campus, School of Economics, University of Nottingham  
Malaysia Campus

May 2018

Online at <https://mpra.ub.uni-muenchen.de/87560/>  
MPRA Paper No. 87560, posted 24 June 2018 16:19 UTC

# **The effects of mediums of instruction on educational- and labor market outcomes: Evidence from Malaysia**

Rasyad A. Parinduri\* and Kian Ong \*\*

## **Abstract**

We examine the effects of having English as a medium of instruction on labor market outcomes later in life. We exploit an exogenous variation in mediums of instruction induced by the government of Malaysia's decision to discontinue English-medium public schools and phase them out with Malay-medium public schools in 1970, which fits a fuzzy regression discontinuity design. We find some evidence that having English as a medium of instruction improves labor market outcomes. We explore some mechanisms through which mediums of instruction matter: We find having English as a medium of instruction improves English proficiency, especially reading and writing skills, and increases educational attainment, which in turn increase earnings and improve employability. The evidence is, however, rather weak, if we use robust data-driven inferences in the regression discontinuity design.

*JEL Classification Codes:* H4, I2, J3, O1

*Keywords:* English, mediums of instruction, regression discontinuity design, labor market outcomes

\*Nottingham University Business School, University of Nottingham Malaysia Campus, Jalan Broga, Semenyih, 43500, Selangor, Malaysia.

Email: rasyad.parinduri@nottingham.edu.my

\*\*School of Economics, University of Nottingham Malaysia Campus, Jalan Broga, Semenyih, 43500, Selangor, Malaysia.

Email: kianhowe.ong@nottingham.edu.my.

## 1. Introduction

Many formerly colonized countries change their mediums of instruction from colonial- to local languages—some switch back and forth between the two, or consider to, as they weigh possibly competing goals of having strong national identity, fast learning, and high returns to schooling—but the literature is mixed on whether mediums of instruction affect earnings and employability. Using local languages as mediums of instruction seems to promote national identity but evidence on its effects on labor market outcomes may depend on the countries' institutions, features of the labor markets, and the reform specifics.<sup>1</sup> Whether the commercial language of a country is the local- or colonial language, the education levels at which the government implement the change, or whether the change is compulsory, for example, may determine the effects of the use of local languages on

---

<sup>1</sup> For example, Aspachs-Bracons et al. (2008), who compare a compulsory language reform in Catalonia with the same but non-compulsory reform in Basque Country, find the reform affects identity in Catalonia but does not in Basque Country; using the same compulsory language reform in Catalonia, Clots-Figueras and Masella (2013) find higher exposure to Catalan language in schools strengthens Catalan feelings. Angrist, Chin, and Godoy (2008), Ramachandran (2017), Seid (2016), Taylor and von Fintel (2016), and Eriksson (2014), which examine the effects on labor outcomes, find using local languages (instead of colonial languages) improves labor outcomes but Angrist and Lavy (1997), Munshi and Rosenzweig (2006), and Shastry (2012) find the opposite.

labor outcomes.<sup>2</sup> The identification methods that the papers use, mostly difference-in-differences and instrumental variable techniques, may matter too.

This paper estimates the causal effects of mediums of instruction on educational- and labor market outcomes using a policy change in Malaysia (English as a medium of instruction was replaced in some schools in 1970 with Malay, the local language), which offers a good and interesting research design. One, the government of Malaysia implemented the change at both primary and secondary schools unlike reforms that many papers in the literature have examined.<sup>3</sup> Two, in Malaysia, English is an important, but not the only, commercial language, which allows us to examine the effects of the change in a labor market in which other languages are also used. Three, the policy change fits a fuzzy regression discontinuity design.<sup>4</sup>

---

<sup>2</sup> Morocco (Angrist and Lavy, 1997) and Puerto Rico (Angrist, Chin and Godoy, 2008) change the medium instruction at secondary schools but Ethiopia (Seid, 2016) and South Africa (Taylor and von Fintel, 2016) at primary schools. The change is compulsory in Puerto Rico (Angrist, Chin and Godoy, 2008) and in Morocco (Angrist and Lavy, 1997) but not in India (Munshi and Rosenzweig, 2006; Shastry, 2012), Ethiopia (Seid, 2016; Ramachandran, 2017), and South Africa (Taylor and von Fintel, 2016; Eriksson, 2014).

<sup>3</sup> Two papers that use similar design, Angrist and Lavy (1997) and Angrist, Chin, and Godoy (2008), find different results.

<sup>4</sup> Angrist and Lavy (1997), Angrist, Chin and Godoy (2008), Ramachandran (2017), Seid (2016), and Eriksson (2014) use difference-in-differences; Taylor and

The switch was unexpected, triggered by an ethnic riot in the aftermath of the 1969 general election in which the two major ethnic groups, Malay and Chinese, clashed (Mariasoosay, 1996). The government then made national unity as a top priority so that on July 11, 1969, Mohamed Khir Johari, the then education minister announced that English in English-medium schools would be phased out and replaced with Malay, which took immediate effect in the following academic year, i.e. January 1970. Moreover, our data allow us to examine some mechanisms through which mediums of instruction may affect labor outcomes, in particular human capital accumulation and better choice of occupations.<sup>5</sup>

To implement the fuzzy regression discontinuity design, in the first stage, we regress an indicator of whether a person had English as a medium of instruction on a polynomial function of year of birth that may discontinue between the 1962 and 1963 birth cohorts (and a set of control variables). In

---

von Fintel (2016), Shastry (2012), and Munshi and Rozensweig (2006) use fixed-effects models; Angrist and Lavy (1997) uses instrumental variable technique.

<sup>5</sup> Language affects earnings via human capital: language of instruction facilitates learning, learning accumulates human capital; human capital earns wages in the labor market. One form of human capital, language proficiency has been studied in relation to earning differential between migrants and natives (Bleakley and Chin 2004; McManus, Gould, and Welch 1983; Chiswick and Miller 1995b; Tainer 1988; Grenier 1984; Chiswick 1991; Kossoudji 1988). The language of instruction also affects earnings via occupations (Shastry, 2012; Munshi and Rozensweig, 2006).

the second stage, we regress an employment outcome on the predicted values of the medium of instruction from the first stage and the same polynomial function.

We use the first and second waves of the Malaysia Family Life Survey conducted by RAND in 1976-1977 and 1988-1989 in West Malaysia. We get information on individuals' medium of instruction at schools from the first wave of the survey; we get labor market outcomes from the second wave.

We find some evidence that having English as a medium of instruction improves labor market outcomes. We explore some mechanisms through which mediums of instruction matter: We find having English as a medium of instruction improves English proficiency, especially reading and writing in English and increases educational attainment, which in turn increase earnings and improve employability. The evidence is, however, rather weak, if we use robust data-driven inferences in the regression discontinuity design.

We proceed as follows: Section 2 describes the policy change; Section 3 presents the empirical strategy and data; Section 4 discusses the results; Section 5 concludes.

## **2. The change in the medium of instruction**

On July 11, 1969, Mohamed Khir Johari, the then education minister announced that English-medium schools will be replaced with Malay as the language of instruction, taking immediate effect in the subsequent academic year i.e. January 1970 (Mariasosay, 1996). This announcement came as drastic and unexpected, following the tragic incident of May 13 1969. 178 were reported dead after a clash between two major ethnicities (Malay and Chinese) after the 1969 election that politicized citizenship, ethnicity and the language of instruction. The election outcome not only led to the worst ethnic riot but also brought the government on the verge of collapse: the prime minister resigned and the sultan (symbolic ruler of the country) declared a state of emergency. Replacing English with Malay as the language of instruction was a drastic measure to foster national unity.

Before 1969, English-medium schools date from British colonial times and coexists with three mediums of instruction: Malay, Chinese and Tamil, reflecting the country's three ethnicities.<sup>6</sup> At the primary level, there are three mediums of instruction: Malay-, Chinese- and English; at the secondary level, there are two mediums of instruction: Malay- and English-medium. Students from non-English-medium primary schools who would

---

<sup>6</sup> Penang Free School, founded in 1816 is the oldest English-medium school in Malaysia and in Southeast Asia.

like to attend English-medium secondary schools attend a ‘remove’ class for one year.

Leading to and even after independence from the British in 1957, the government through committees and publications (the Barnes Report 1951, the Fenn-Wu Report 1951, the Razak Report 1956, the Education Ordinance 1957, the Rahman Talib Report 1960 and the Education Act 1961) guarantee English-medium schools’ place schools. Therefore, the phasing out of English-medium schools was unforeseen before 1969. In 1970, the option of English-medium school is removed from the ‘education menu’ available to parents and students. For the purpose of our identification, the treatment i.e. English as language of instruction is switched off in 1970.

The switch from English to Malay as the language of instruction occurs gradually that suits regression discontinuity design. The switch begins at the primary schools in 1970 and completes in 1975. Subsequently, the switch begins at the secondary schools in 1976 and completes in 1982 (the second Malaysia Plan 1971-1975; the third Malaysia Plan 1976-1980 (para: 1310, 1364)). The Fourth Malaysia Plan 1981-1985 (para: 970, 998) confirms the completion in 1982 in West Malaysia (Our dataset covers the West Malaysia) that is consistent with six years of primary education and six years of secondary education (Malaysia: Mid Term Review of the Fourth Malaysia Plan 1981-1985, para 842).

Several caveats might distort our study: first, does the language switch affect the quality of education due to teacher training or (shortage of) resources? Second, does the language switch alter the commercial language of the labor market concurrently? To the first caveat, due to the unplanned nature of the policy, schools are reportedly short of teachers on subjects in Malay, only to be remedied by lengthening the operation hours of training schools. The Language Institute and other teaching training colleges coordinate the training and retraining of teachers.<sup>7</sup> Dewan Bahasa and Pustaka coordinates the provision of textbooks (Mariasoosay, 1996). Mariasoosay (1996) noted that the influence on quality of education is inevitable. To the second caveat, the language switch affects the language used in the public sector but not in the private sector. English remains the commercial language in the private sector (Mariasoosay, 1996).

---

<sup>7</sup> The government trained 9274 primary and secondary school teachers in 1971-1973; additional 4,540 teachers with intensive training in Bahasa Malaysia during 1974-1975 was envisaged. (Mid Term Review of the second Malaysia Plan 1971-1975, para: 585).

### **3. Empirical strategy and data**

#### *3.1. Empirical strategy*

This policy change fits a regression discontinuity (RD) design, where using English as medium of instruction is our variable of treatment. English is discontinued as medium of instruction in 1970 onwards, we can exploit this policy change as a source of exogenous variation in English. The source of exogeneity can be determined by year of birth, where we define as the running variable. The idea of the RD is, by employing the exogenous variation of treatment variable and the source of exogeneity, around the discontinuity, we can estimate the causal effects of the treatment on outcome variables of interest (see Lee and Lemieux, 2010). People who were born in 1962 or earlier and entered primary school in 1969 or earlier might attend an English-medium public school; people who were born in 1963 or later and entered primary school in 1970 or later attended an English-medium school only if they attended an international school, which very few Malaysians did. There is therefore a discontinuity in the probability of attending an English-medium school between the 1962 and 1963 birth cohorts, which we use as an instrumental variable to generate an exogenous variation in mediums of instruction in a regression discontinuity design.

The RD design is fuzzy because the treatment variable, English is not a deterministic function of the running variable, year of birth.

Subsequent to policy change, the probability of receiving English-medium school is not zero as individuals could attend private schools. It could also be that some individuals attend school a year earlier or later than the official age of seven when entering primary school. We use a fuzzy RD, that entails a two-stage least square (2SLS) estimation, where the year of birth is used as instrument for English to estimate its causal effects on outcome variables of interest.

We describe the first and second stage of the 2SLS in turn. In the first stage we predict the proportion of individuals who have English as medium of instruction before and after the policy change:

$$English_i = \alpha + \beta OldCohorts_i + f(yob_i) + \epsilon_{1i} \quad (1)$$

$English_i$  is defined as using English as medium of instruction which is our variable of treatment;  $OldCohorts_i$  identifies the discontinuity in probability of treatment when the policy change occurs:

$$OldCohorts_i = \begin{cases} 1, & yob_i \leq 1962 \\ 0, & yob_i > 1962 \end{cases}$$

$yob_i$ , year of birth is the running variable,  $f(yob_i)$  are other relevant predictors of the probability of treatment which include dummy variables for ethnicity, a cubic polynomial function of age and an indicator for attending remove class. In the second stage, the predicted value of English obtained from the first stage,  $\widehat{English}_i$  is inserted as a control variable, alongside other controls from the first stage:

$$wages_i = c + \lambda (\widehat{English}_i) + f(yob_i) + \epsilon_{2i} \quad (2)$$

$wages_i$  is the monthly wages;  $\lambda$  is the causal estimate of English as medium of instruction on wages.

We also use robust data-driven inferences in the RD design by estimating both the reduced-form (that is, sharp RD) and the two-stage least-squares (that is, fuzzy RD) of the following RD equation:

$$y_i = \alpha + \beta D_i + f(mob_i) + \varepsilon_i \quad \forall x_i \in (c - h, c + h) \quad (3)$$

where  $y_i$  is an outcome variable,  $D_i$  is an indicator equals one if person  $i$  is a member of the older cohorts;  $f(mob_i)$  is a polynomial function of the assignment variable  $mob_i$ , the month of birth; and  $h$  is an optimal bandwidth around the cutoff point  $c$ , which we estimate using Imbens and Kalyanaraman's (2012) algorithm. We implement local linear regressions of Equation (1) using Calonico, Cattaneo and Titiunik's (2017) software for RD designs.

### 3.2. Data

We use the first and second waves of the Malaysia Family Life Survey, a RAND's survey in peninsular Malaysia in 1976-1977 and 1988-1989. The first wave has 1,262 households; the second 926. About 3,900 individual adults were interviewed in the second wave. We get the information on mediums of instruction people had when they were in school

from the first wave; we get the labor outcomes when they were in the labor market from the second wave.

The outcome of the first-stage regressions, *English*, is a dummy variable equals one if somebody had English as the medium of instruction and zero otherwise. The outcome of the second stage is monthly wages in RM 1988. We also include gender, a set dummy variables for races, and a cubic function of age as control variables in some specifications.

## 4. Results

### 4.1. Basic results

Figure 1, which illustrates the first-stage regression, shows that the trend line of the proportion of people who had English as the medium of instruction breaks between the 1962 and 1963 birth cohorts: We can therefore use *OldCohorts*, a dummy variable equals one if somebody was born in 1962 or earlier, as an instrumental variable for having English as the medium of instruction in a fuzzy RD design. One in five the early 1950s cohorts had English as the medium of instruction, as many as two in five for the late 1950s, and about one in five for the 1962 birth cohorts. But much fewer people in the 1963 birth cohorts did; those who did probably entered primary school when they were six years old in 1970. Between the 1962 and

1963 birth cohorts, the proportion of people who had English as the medium of instruction falls about 10-15 percent.

<Insert Figure 1 here>

The estimates in Panel A of Table 1 confirm the break in the trend line. The estimates vary from 0.23 when we use a linear function of year of birth as control to 0.12 when we use cubic polynomial function of year of birth, cubic polynomial function of age, gender and a set of dummies for races as control. Compared to people in the 1963 birth cohorts, people in the 1962 birth cohorts were about 10 percent more likely to have English as the medium of instruction.

<Insert Table 1 here>

Figure 2, which illustrates the second-stage regression, shows that there is some evidence that the trend line of monthly wages break between the 1962 and 1963 birth cohorts. The data are noise as the averages show, especially for early 1950s cohorts, most probably because we have fewer number of observations for the earlier cohorts. But, we still see that the average monthly wages of the early 1960s birth cohorts are about RM 500-600 while that of the 1963 birth cohort about RM400. The trend line indicates that there is a fall of about RM100 or more between the 1962 and 1963 birth cohorts.

<Insert Figure 2 here>

The estimates in Panel B of Table 1 show that monthly wages do drop between the 1962-1963 birth cohorts. When we use a linear function of year of birth as control, the estimate is about RM500; when we include all sets of control RM1,300—a large estimate considering that the average monthly wages at the time is about RM500.

#### *4.2. Mechanisms*

We explore three mechanisms through which English as a medium of instruction may affect labor outcomes later in life: English proficiency, educational attainment, and having better jobs.

Figure 3, which illustrate the pathway through English proficiency, shows that having English as a medium of instruction may affect reading and writing proficiency in English, but perhaps not speaking English. There are slight drops between the 1962 and 1963 birth cohorts in Figure 3 (a) and (b), but none in Figure 3 (a).

<Insert Figure 3 here>

Figure 4, which illustrate the pathway through educational attainment, shows that there is a drop in average educational attainment between the 1962 and 1963 birth cohorts. The drop is small but, given that it is driven by only a ten percent increase in the probability of having English as the medium of instruction, the magnitude of the effects of having English as the medium of instruction on educational attainment is large.

<Insert Figure 4 here>

Figure 5, which illustrate the pathway through getting better jobs, shows there is also a drop in the probability of securing white collar jobs between the 1962 and 1963 birth cohorts. Again the drop is small but economically it implies a large effect of having English as the medium of instruction.

<Insert Figure 5 here>

Table 2, which presents the estimates, confirms the drops in Figures 3-5: There is some evidence that having English as the medium of instruction affects labor outcomes later in life through an improvement in English proficiency, and an increase in educational attainment, and an increase in the probability of securing white-collar jobs (Panel A). The estimates in column (2) show that having English as the medium of instruction increases the likelihood of proficient in reading and writing in English by 53 and 45 percent, respectively, though they are statistically significant at ten percent level. Having English as the medium of instruction also increases the probability of proficient in speaking English but the estimate is statistically insignificant.

<Insert Table 2 here>

Having English as the medium of instruction also seems to increase educational attainment and the probability of securing better jobs. The drop

in the trend line in Figure 4 is small, but the estimate of the effect on educational attainment is large, 3-4 years (Panel B of Table 2). Having English as the medium of instruction also increases the likelihood of securing better jobs by about 50 percent or more (Panel C).

#### *4.3. Robust data-driven inferences*

Using robust data-driven inferences in the RD design in which we use month of birth as the assignment variable and the optimal bandwidths are calculated using Imbens and Kalyanaraman's (2012) algorithm, overall we find similar results in terms of the signs of the estimates. We still see the discontinuity in the probability of having English as the medium of instruction (as Panel (a) of Figure 6 illustrates) that is statistically significant, but the estimates of the effects of the medium of instruction on wages are statistically insignificant (Panel (b)); the same applies to the mechanisms that we consider such as educational attainment, English proficiency, or job characteristics (Panel (c)).<sup>8</sup>

<Insert Figure 6 here>

---

<sup>8</sup> We do not present the estimates of the sake of brevity; they are available from the authors upon request.

#### **4. Concluding remarks**

We examine the effects of having English as a medium of instruction on education and labor outcomes later in life using the first and second waves of the Malaysia Family Life Survey. We exploit an exogenous variation in mediums of instruction induced by the government of Malaysia's decision to discontinue English-medium public schools and phase them out with Malay-medium public schools in 1970, which fits a fuzzy regression discontinuity design. People who were born in 1962 or earlier and entered primary school in 1969 or earlier might attend an English-medium public school; people who were born in 1963 or later and entered primary school in 1970 or later attended an English-medium school only if they attended an international school, which very few Malaysians did.

There is therefore a discontinuity in the probability of attending an English-medium school between the 1962 and 1963 birth cohorts, which we use as an instrumental variable to generate an exogenous variation in mediums of instruction in a regression discontinuity design. To implement the fuzzy regression discontinuity design, in the first stage, we regress an indicator of whether a person had English as a medium of instruction on a polynomial function of year of birth that may discontinue between the 1962 and 1963 birth cohorts (and a set of control variables). In the second stage, we regress an employment outcome on the predicted values of the medium of instruction from the first stage and the same polynomial function.

We find some evidence that having English as a medium of instruction improves employment outcomes though it is rather weak if we use robust data-driven inferences. We explore some mechanisms through which mediums of instruction matter: We find having English as a medium of instruction improves English proficiency, especially reading and writing in English; it also increases educational attainment—these in turn increase earnings and improve other employment outcomes later in life.

## **References**

- Angrist, J., Chin, A., and Godoy, R. 2008. "Is Spanish-Only Schooling Responsible for the Puerto Rican Language Gap?" *Journal of Development Economics* 85 (1): 105-128.
- Angrist, J., and Lavy, V. 1997. "The Effect of a Change in Language of Instruction on the Returns to Schooling in Morocco." *Journal of Labor Economics* 15 (1, Part 2): S48-S76.
- Aspachs-Bracons, O. , Clots-Figueras, I. , Costa-Font, J., and Masella, P. 2008. "Compulsory Language Educational Policies and Identity Formation." *Journal of the European Economic Association* 6: 434-444.
- Bleakley, H., and Chin, A. 2004. "Language Skills and Earnings: Evidence from Childhood Immigrants." *Review of Economics and Statistics* 86 (2): 481-496.
- Calonico, S., Cattaneo, M.D., and Titiunik, R. 2015. "Optimal data-driven regression discontinuity plots." *Journal of the American Statistical Association* 110(512): 1753-1769.
- Chiswick, B.R. 1991. "Speaking, Reading, and Earnings among Low-Skilled Immigrants." *Journal of Labor Economics* 9 (2): 149-170.
- Chiswick, B.R., and Miller, P.W. 1995. "The Endogeneity between Language and Earnings: International Analyses." *Journal of Labor Economics* 13 (2): 246-288.
- Clots-Figueras, I. and Masella, P. 2013. "Education, Language and

- Identity." *Economic Journal*, 123: F332-F357.
- Eriksson, K. 2014. "Does the Language of Instruction in Primary School Affect Later Labour Market Outcomes? Evidence from South Africa."
- Economic History of Developing Regions* 29 (2): 311-335.
- Grenier, G. 1984. "The Effects of Language Characteristics on the Wages of Hispanic-American Males." *The Journal of Human Resources* 19(1): 35-52.
- Imbens, G.W., and Kalyanaraman, K. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *Review of Economic Studies* 79(3): 933-959.
- Kossoudji, S.A. 1988. "English Language Ability and the Labor Market Opportunities of Hispanic and East Asian Immigrant Men." *Journal of Labor Economics* 6 (2): 205-228.
- Lee, D.S., and Lemieux, T. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48(2): 281-355.
- Mariasoosay, T. 1996. *Perspectives on language policies in Malaysia*, Durham theses, Durham University. Available at Durham E-Theses Online: <http://etheses.dur.ac.uk/5185/>
- McManus, W., Gould, W., and Welch, F. 1983. "Earnings of Hispanic Men : The Role of English Language Proficiency." *Journal of Labor Economics* 1(2): 101-130.
- Munshi, K., and Rosenzweig, M. 2006. "Traditional Institutions Meet the

- Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy.” *American Economic Review* 96(4): 1225-1252.
- Ramachandran, R. 2017. “Language Use in Education and Human Capital Formation: Evidence from the Ethiopian Educational Reform.” *World Development* 98: 195-213.
- Seid, Y. 2016. “Does Learning in Mother Tongue Matter? Evidence from a Natural Experiment in Ethiopia.” *Economics of Education Review* 55: 21-38.
- Shastry, G.K. 2012. “Human Capital Response to Globalization.” *Journal of Human Resources* 47 (2): 287-330.
- Tainer, E. 1988. “English Language Proficiency and the Determination of Earnings among Foreign-Born Men.” *Journal of Human Resources* 23 (1): 108-122.
- Taylor, S., and von Fintel, M. 2016. “Estimating the Impact of Language of Instruction in South African Primary Schools: A Fixed Effects Approach.” *Economics of Education Review* 50: 75-89.

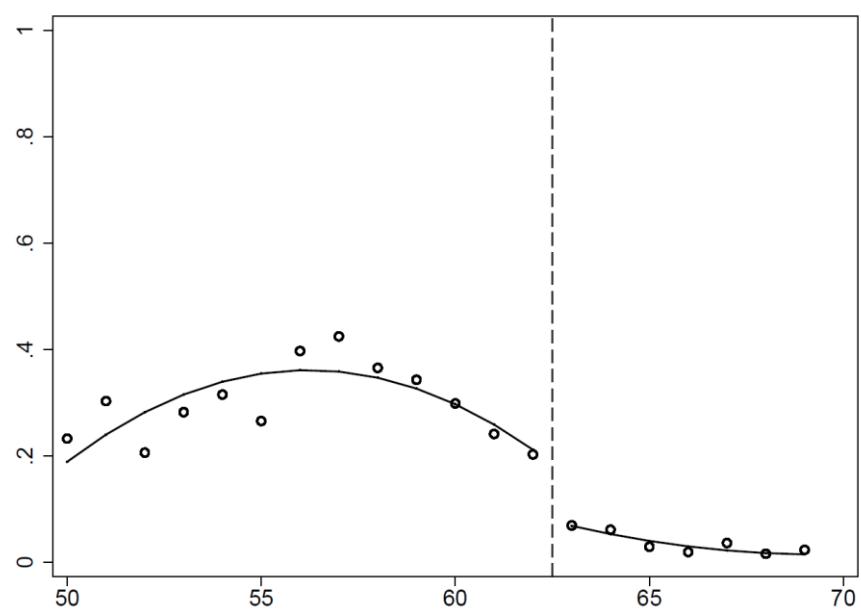


Figure 1 Had English as the medium of instruction

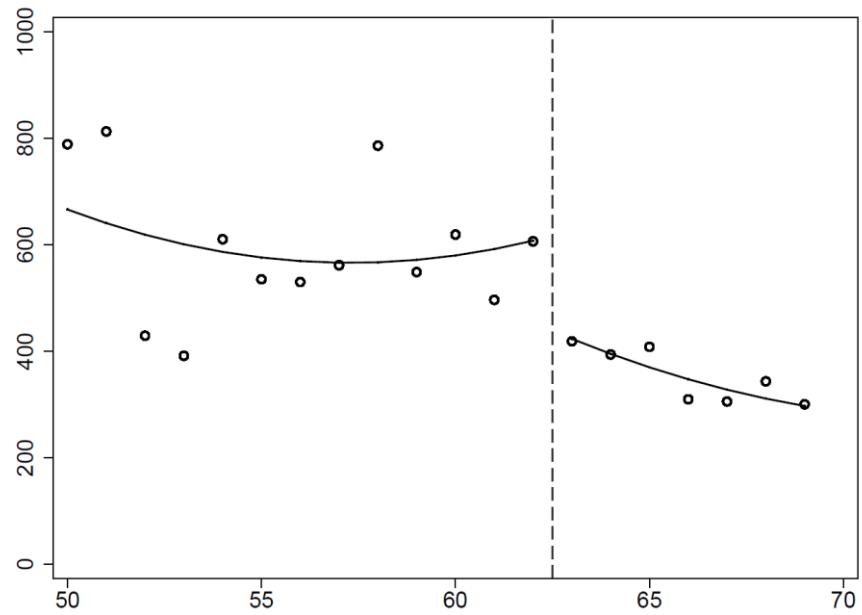
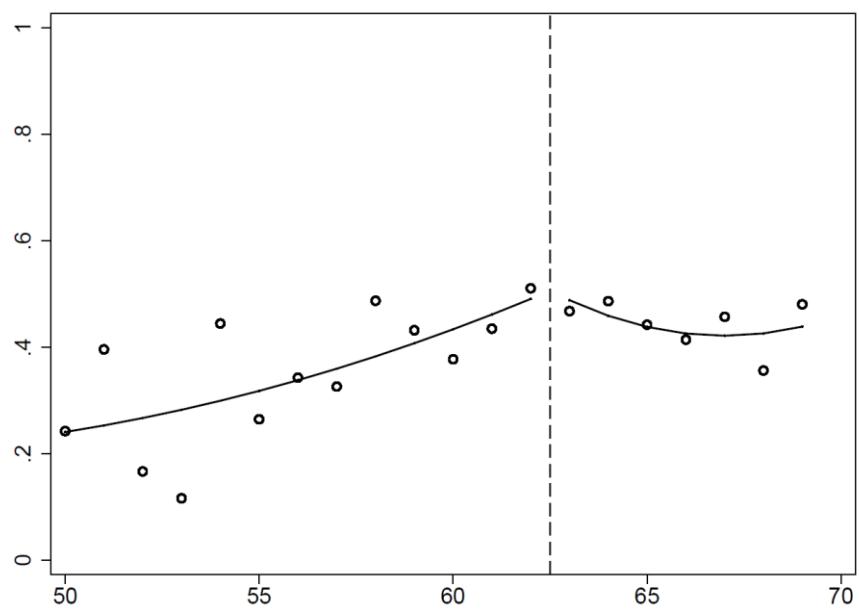
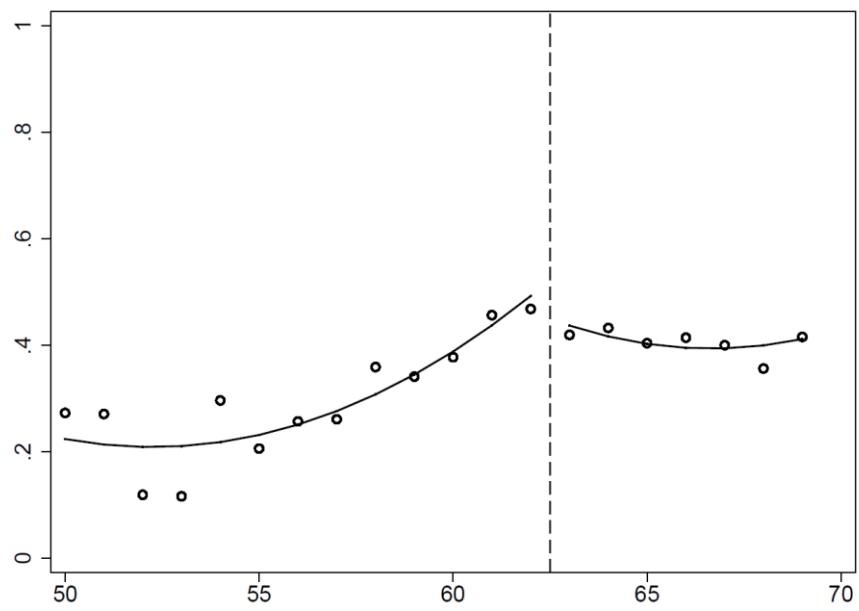


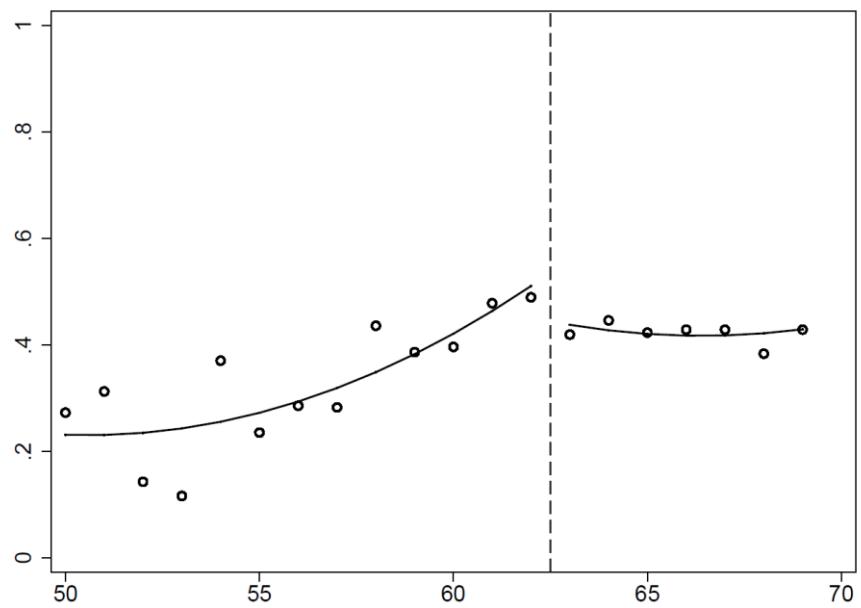
Figure 2 Monthly wages (RM 1986)



(a) Able to speak English



(b) Able to read in English



(c) Able to write in English

Figure 3 English proficiency

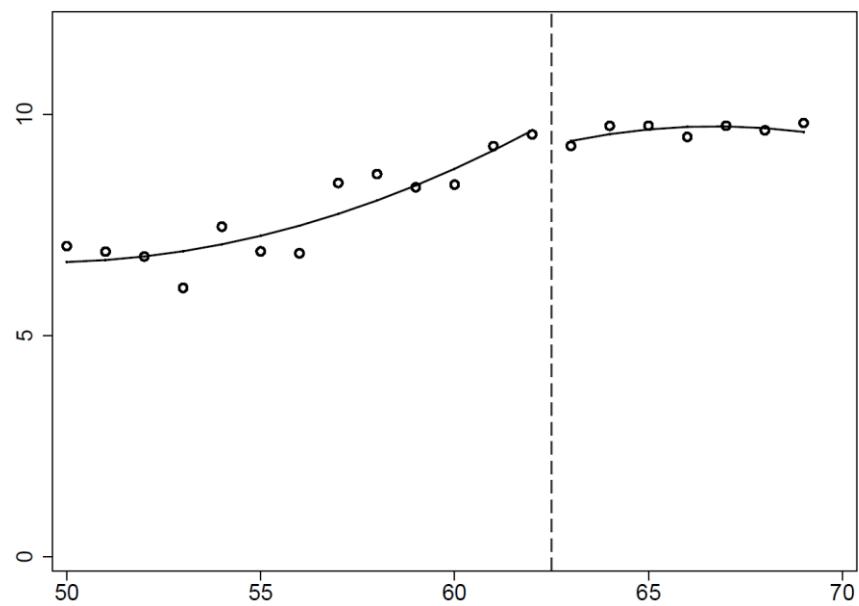


Figure 4 Educational attainment

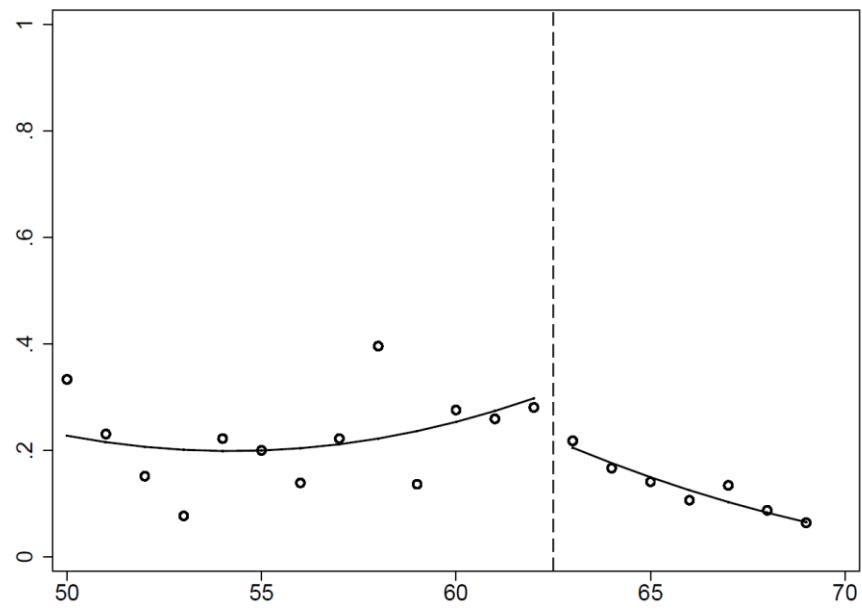
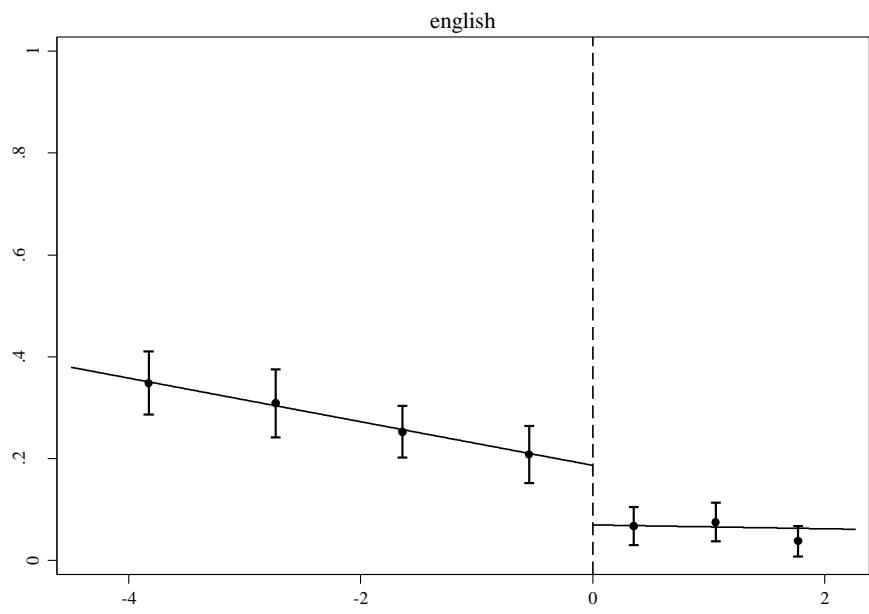
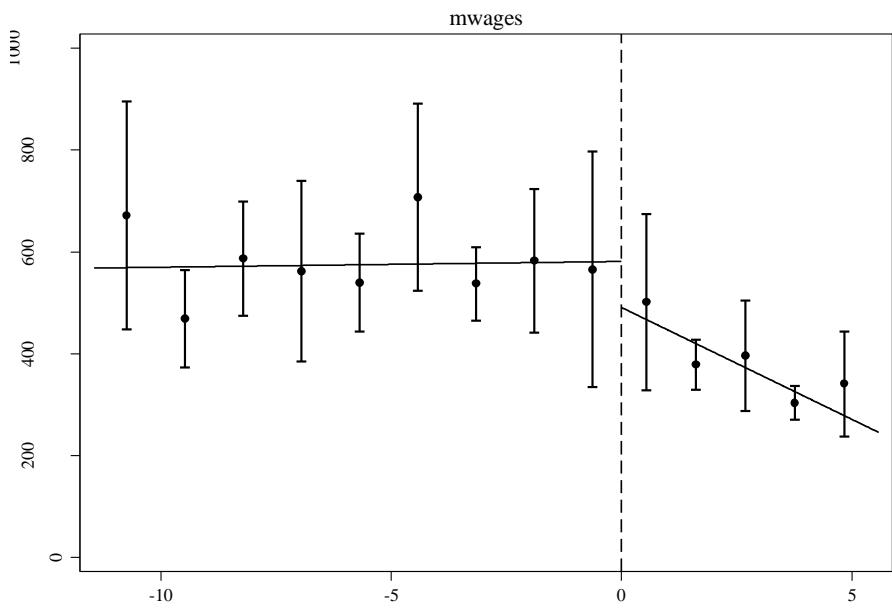


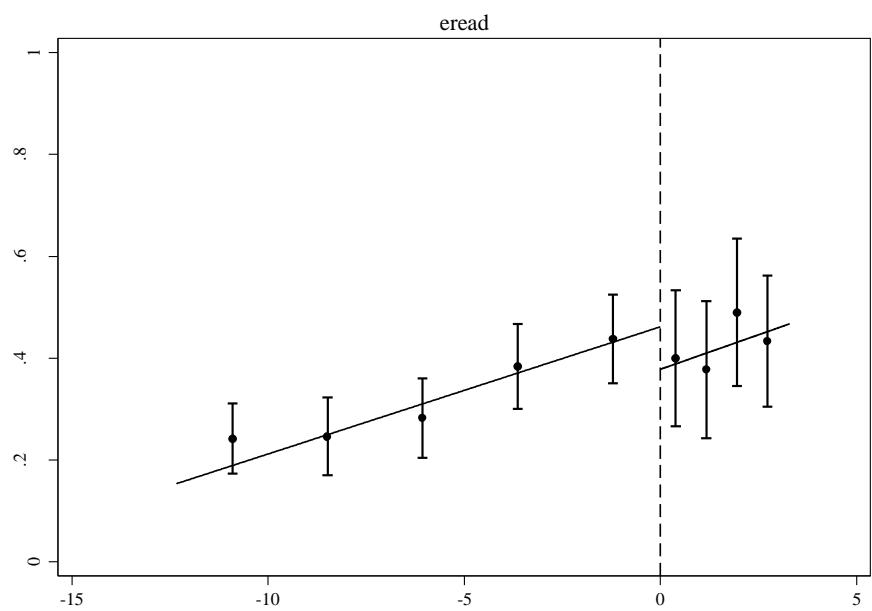
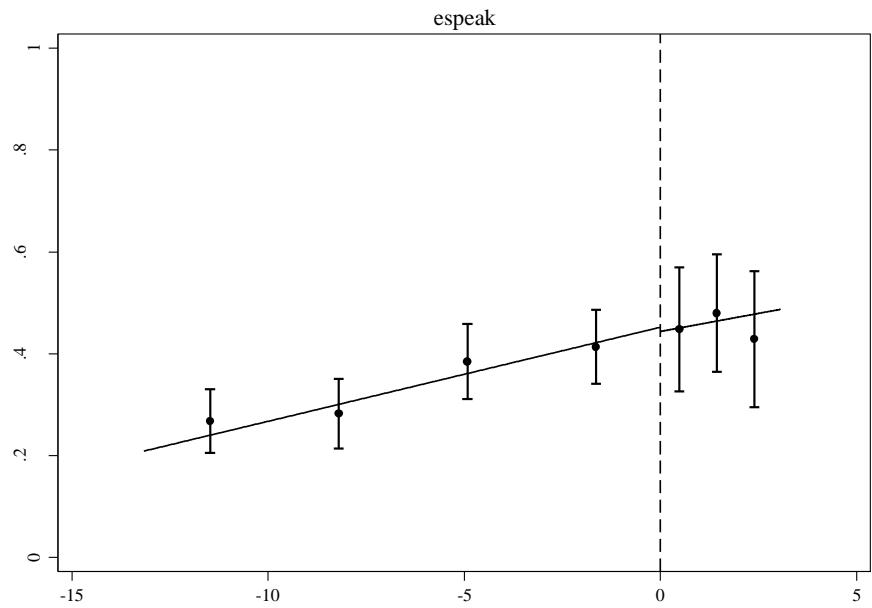
Figure 5 Had a white-collar job

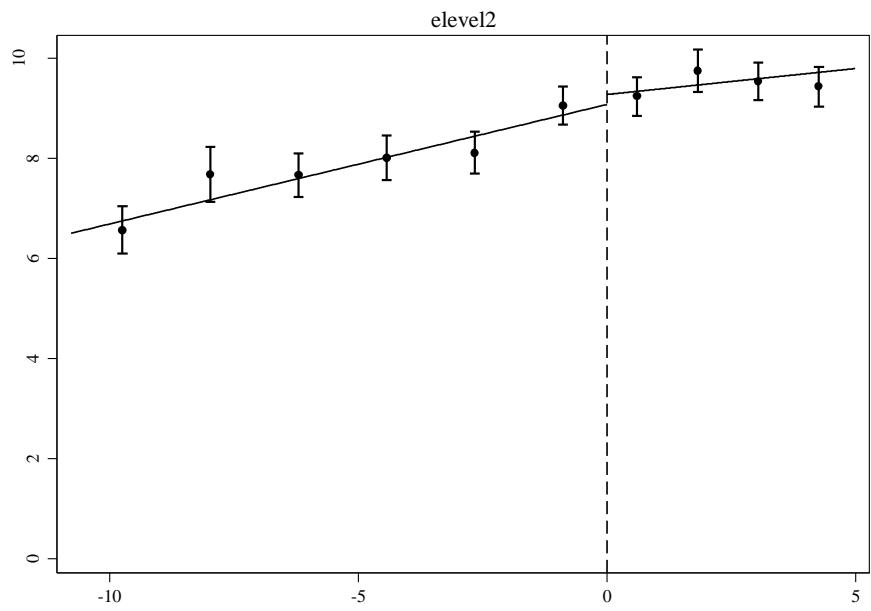
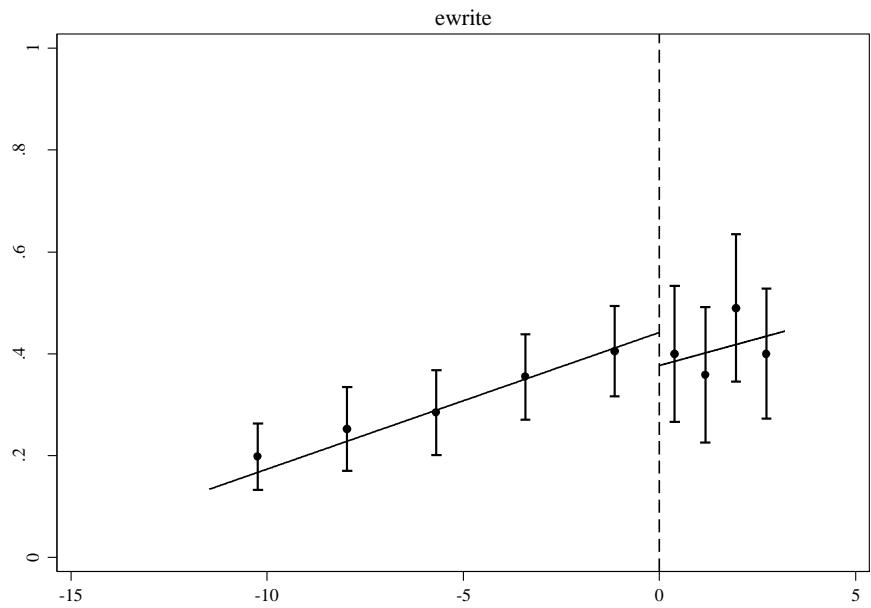


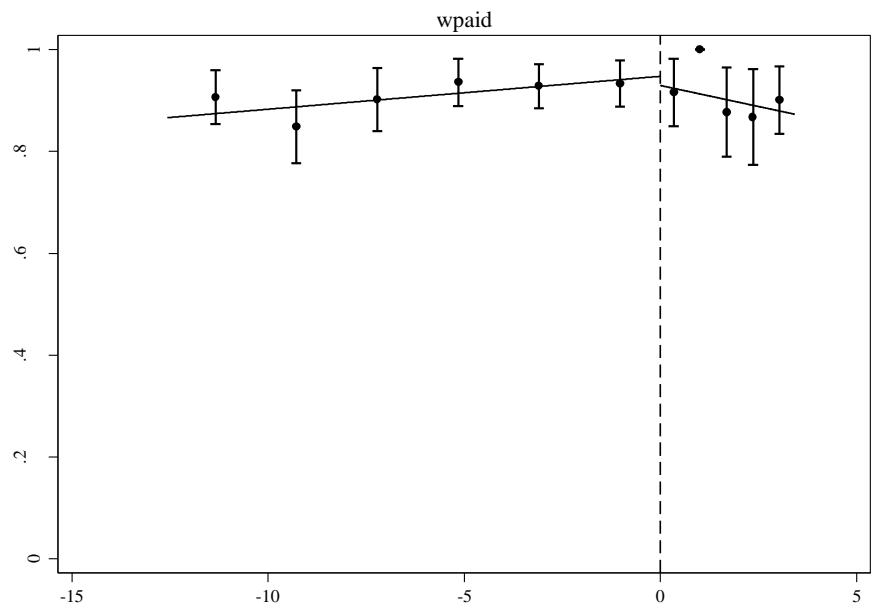
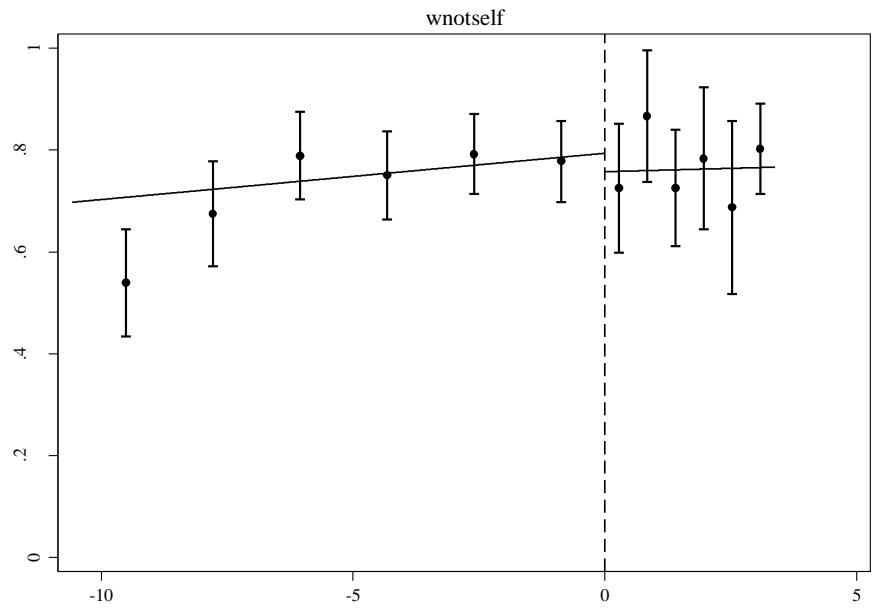
(a) Had English as a medium of instruction

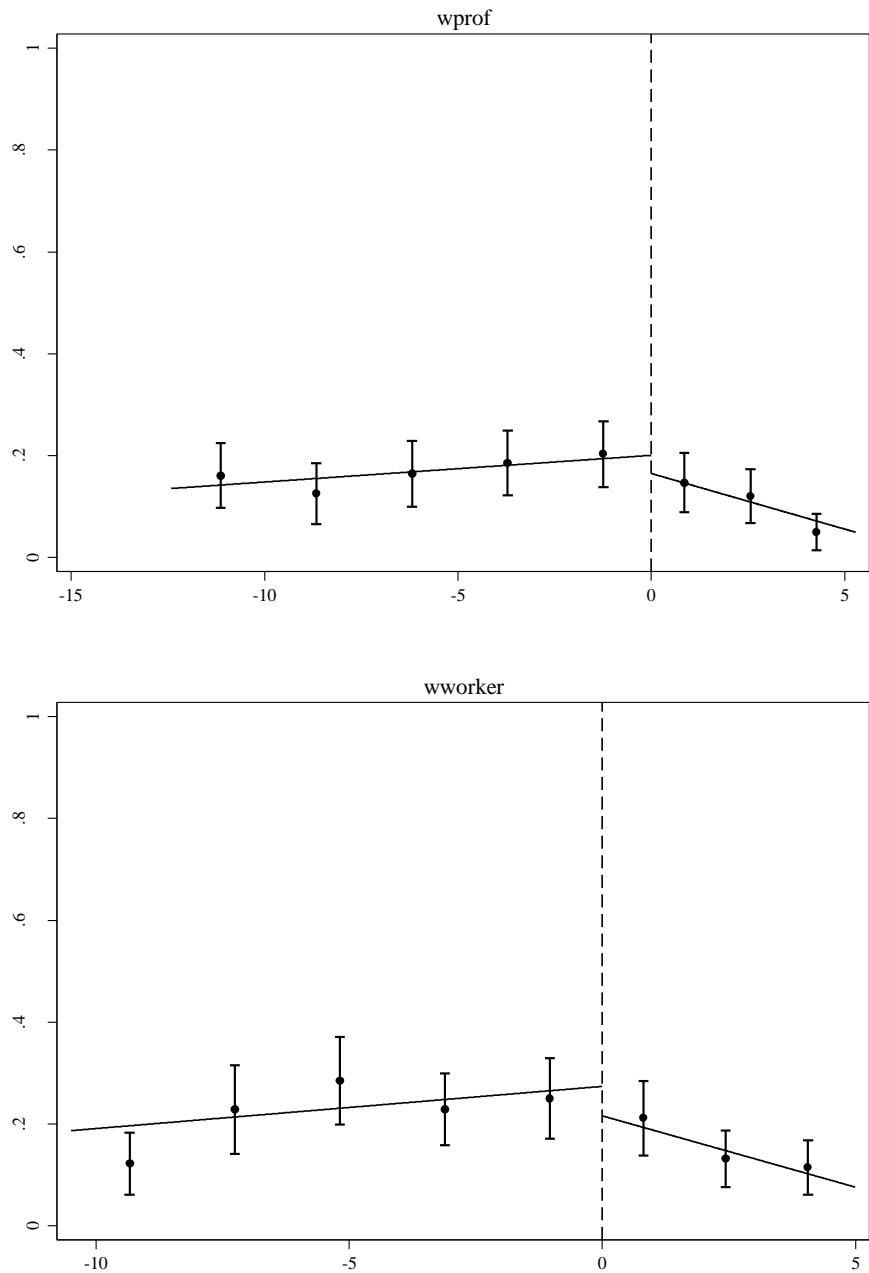


(b) Wages









(c) Mechanisms

Figure 6 Robust data-driven inferences in the RD design

Table 1 The first- and second-stage results

	(1)	(2)	(3)	(4)	(5)
<b>A. First-stage regression</b>					
Dependent variable: Had English as a medium of instruction					
Older cohorts	0.23 (0.02)	0.21 (0.02)	0.13 (0.03)	0.13 (0.04)	0.12 (0.04)
Adjusted R <sup>2</sup>	0.13	0.13	0.14	0.12	0.17
Number of observations	3,902	3,902	3,902	1,862	1,857
<b>B. Second-stage regression</b>					
Dependent variable: Monthly wages					
Had English as a medium of instruction	556 (264)	593 (283)	1,181 (601)	1,194 (604)	1,309 (708)
Number of observations	1,045	1,045	1,045	1,045	1,042
Controls					
<i>Year of birth</i>		✓			
<i>Year-of-birth quadratic polynomial</i>			✓		
<i>Year-of-birth cubic polynomial</i>				✓	✓
<i>Age cubic polynomial</i>				✓	✓
<i>Other control variables</i>					✓

Notes: In Panel A, the number in each cell is the estimate of older cohorts from a separate regression of whether a person had English as a medium of instruction on older cohorts and a set of control variables. In Panel B, the number in each cell is 2SLS estimate of the effects of having English as a medium of instruction on monthly wages (in RM 1986). Older cohorts equals one if an individual was born in 1962 or earlier and zero otherwise. The figures in parentheses are robust standard errors.

Table 2 Mechanisms

	(1)	(2)
A. English proficiency		
1. Speaking	0.47 (0.51)	0.28 (0.29)
2. Reading	1.00 (0.58)	0.53 (0.28)
3. Reading	0.93 (0.56)	0.45 (0.28)
B. Educational attainment	4.14 (2.33)	3.32 (1.84)
C. White-collar jobs	0.71 (0.37)	0.47 (0.21)
Polynomial function of year of birth		
<i>Cubic function</i>	✓	
<i>Cubic function that may differ before and after 1962</i>		✓

Notes: The number in each cell is the 2SLS estimate of having English as a medium of instruction on a dependent variable listed on the left column. The figures in parentheses are robust standard errors.

