ERF 2020

26TH Annual Conference

Effect of English-Medium Instruction on Earnings

Murat Tasdemir & Yılmaz Kılıçaslan

ECONOMIC RESEARCH FORUM
منتدى البحوث الاقتصادية

Effect of English-Medium Instruction on Earnings*

Murat Taşdemir†

Yılmaz Kılıçaslan‡

25 Kasım 2019

Abstract

The adoption of English-medium instruction in high schools and universities has been controversial in non-English speaking countries including Turkey. Research on the effect of medium of instruction in higher education carries importance in term of both educational policy making and school choice of individuals. This is the first study to examine the effect of English medium higher education programs on earnings in as setting where English is neither dominant not a colonial language. Using a data set from an online survey, we carry out econometric estimations that take into account possible selections problems. The results suggest that graduating from an English-medium business or economics program in Turkey raise monthly earnings by 11%-22% compared to a Turkish-medium program in the same major.

1 Introduction

The effect of graduating from an undergraduate program with English-medium instruction on labor market outcomes is of a great importance for both educational policies and individual decisions. The adoption of English-medium instruction in high schools and universities has been controversial in non-English speaking countries including Turkey (Coleman, 2006; Ekoç, 2018; Kılıçkaya, 2006; Selvi, 2014). Turkey makes an interesting case because it is a monolingual country, and has no colonial history with European nations. Proponents of English medium instruction in higher education base their arguments mostly on pedegogical

*This work is funded by Turkish Scientific and Technological Research Council (TUBITAK) with the project no. 115R305

†İstanbul Medeniyet University (corresponding author).

‡Anadolu University.

effectiveness and economic grounds (Özen et al., 2014; Selvi, 2014). Economic arguments in favor of English medium instruction in higher institutions are namely social and private benefits of English language skills. Argued private benefits of graduating an English medium program (EMP) are mainly higher wages and better career opportunities in labor market. They argue that English is important for the development process of the country through promoting globalization, and welfare of individuals through higher salaries and better career opportunities (Özen et al., 2014). On the other hand, the empirical evidence in this realm is very limited. We do not really know that English-medium instruction in higher education makes any difference in labor market outcomes.

Evidence-based policymaking with regard to increasing human capital stock in the economy requires the knowledge on the possible effect of alternative policies. Introduction of English-medium undergraduate programs has important implications in terms of efficient allocation of national resources, as well as individual career choices. If these programs are shown to have a negligible or even insignificant effect on labor market outcomes, then they can be regarded as just a tool of attracting high ability students for universities. Furthermore, these programs may be detrimental to human capital stock of the country, since there is some empirical evidence that in some context, foreign language instruction hampers learning (Hu & Alsagoff, 2010; Kırkgöz, 2014; Selvi, 2014). However, if their effect is shown to be positive and significant, then extending English-medium programs will lead to creating internationally competitive and more productive labor force. Therefore, any research on the extent to which graduating from an undergraduate program offered in English affects earnigs in developing countries, such as Turkey carries importance. This is the first study to examine the effect of English medium higher education programs on earnings in as setting where English is neither dominant not a colonial language. The data set used in this study allows to properly estimate this effect.

The effect of graduating from an English-medium undergraduate program on earnings may arise through two different but possibly simultanous mechanisms. The first is through increased human capital, and in turn higher labor productivity. Those who graduate from

English-medium programs may accumulate higher human capital that raise their productivity in the labor market, hence are expected to receive higher earnings compared to the graduates from Turkish-medium programs. The difference human capital accumulation may result from foreign language skills, as well as from the differences in their knowledge and skills related to their fields of study. The facilitation of the communication in foreign languages may enhance marginal productivity of employees, and encourage the employers to pay higher wages in the form of language skill premium. This can also allow employees who speak foreign languages to earn even more in the case of a shortage of employees with such qualifications. These employees can have better careers, which will raise their earnings. The difference in the human capital between English-medium and Turkish-medium graduates may also arise from the skills related to their field of study. English-medium programs may have a better academic curriculum. In addition, the teaching staff in English-medium programs may be, on average, more qualified, so that students in these programs acquire more field knowledge.

The second mechanism through which graduating from an English-medium program could affect labor market outcomes is through the "signaling". In Turkish higher education system, students take a nationwide exam and get assigned to programs by the ÖSYM (Student Selection and Placement Center) according to students' exam scores and their states preferences among programs. Suppose that the English language skills have no return in the economy and there is no difference between the programs in terms of learning field knowledge. Enrolling to English-medium programs requires higher scores in the entrance exam than Turkish-medium programs in the same major. High type students will have lower opportunity costs related to their enrolment in such Englihs-medium programs. Students who want to signal that they are more productive are more likely to choose English-medium programs. Hence, graduating from an English-medium program will send a signal to employers that they are more productive, independent of their English proficiency. Consequently, English-medium graduates will earn more than the Turkish-medium graduates in the same major. The data set used in this study, does not allow us to decompose the human capital resultinf from English proficiency or field knowledge, and signalling effects of English-medium instruction in undergraduate programs.

It would be misleading to interpret the findings of this study as the English language premium.

As in many econometric research with observed data, estimations of the effect of graduating from English-medium undergraduate programs on earnings may be biased due to endogeneity problems. Endogeneity arises from the fact that students are not assigned to English or Turkish-medium programs randomly. If students who chose and got assigned to English-medium programs are more talented on average, after graduation, they will work for higher wages compared to the graduates from a Turkish-medium program in the same major, regardless of the medium of instruction in the program they graduated. Therefore, it cannot be identified whether the estimated coefficients reflect the ability of students, or their skills acquired by studying in an English-medium program. Any estimation not taking this identification problem will be subject to selection bias. In this paper, we attempt to control possible selection bias resulting from two different sources, namely sample selection and endogenous selection (self-selection).

This is the first study looking at the effect of foreign language instruction in higher educations on earnings. Although there are many studies analyzing the returns to foreign language skills, most of them focus only on the effect of skills in a dominant local language on the earnings of immigrants in developed countries. There are few number of sudies on the returns to speaking a foreign language. Studies in the literature, on the effect of the language of instruction in educational institutions on earnings are limited to primary and secondary school environments.

2 Related Literature

Although there is a large liteature on returns to language skills, the bulk of the papers in the subject are concerned with the value of the native language to immigrants in labor markets. Most of these papers are on the return to English skills of Spanish-speaking immigrants to the United States (see Bleakley & Chin (2004) for a brief review). There are few studies on the value of a foreign language in environment that is not the dominant language, and yet fewer studies that focus on the effects of a foreign language as a medium of instruction in schools.

Among those who look at the returns to English language skills, Levinsohn (2004) estimates the returns to English in South-Africa, Munshi & Rosenzweig (2006) in India and Lang & Siniver (2006) in Israel. All find significantly positive returns to English language proficiency with one exception that, in South-Africa, it is true only for white people.

Estimating the effect of the instructional language on labor market outcomes is not the same as estimating the effect of language proficiency. In the first case, two possible mechanism may act simultaneously. The first is the true effect of the language skills acquired by the language of instruction. The language of instruction may affect earnings or other labor market outcomes by increasing the language skills, and in turn labor productivity. The other effect works through signaling the employer that the individual is a high productivity type. This latter effect is expected to be prevalent where the medium of instruction is chosen by indiviuals. The choice of instructional language is generally realized by the school choice. The studies by Angrist & Lavy (1997), Angrist, Chin, & Godoy (2008), Chakraborty & Bakshi (2012) and Anghel, Cabrales, & Carro (2012) look at the effect of the instructional language on labor market outcomes. All these studies make use of some governmental program that changes the language of instruction in primary or elementary schools.

An earlier work on the effect of the language of instruction on labor market outcomes is the paper by Angrist & Lavy (1997). By employing a *difference in differences* strategy, Angrist & Lavy (1997) estimate the effect of the change in the instructional language used in primary schools in Morocco after 1983 from French to Arabic on *writing skills* and *earnings*. French being the dominant language in Morocco and the region, plays the same role as English in the Turkish labor market today. They use two sources of data: Living Standards Measurement and Literary Survey (LSMS) conducted in 1991, and Moroccan Labor Force Survey (LFS) conducted in 1990 and 1991. The authors identify the individuals who were exposed to the change in the language of instruction by their ages. Those who are younger than 21 were assumed to be exposed to "Arabization" program.

Using both data sets, the authors estimate Mincerian wage equations, and find that the change in the language of instruction has significant negative effect on earnings. Moreover, they estimate the effect of the program on French writing skills, and conclude that it has also a negative effect on French writing skills. Therefore, one possibility is that the change of instruction from French to Arabic reduce earnings of Moroccan young through French writing skills. In their case, the effect of the change in the instructional language works through the individual's French language skills. In order to identify the direct effect of the program, the authors also perform instrumental variable estimations with French writing skills index variable as an explanatory variable. The estimation results confirm that the change in the language of instruction had an effect on earnings. In Angrist & Lavy (1997), there is no selection problem because the change occurs exogenously. It is implemented in the whole country. Since there is no cross-sectional variation their data, they face an identification problem. They basically compare the earnings of those before and after the government's program. In such a setting, if there is an increase in the French language premium over time, then the estimations would be biased. In order to identify the true effect they employ the difference in differences approach. Their work is still subject to ability bias though.

Two other studies along the same lines are Angrist et al. (2008) and Chakraborty & Bakshi (2012). Angrist et al. (2008) investigates the effect of Puerto Rican change in the language of instruction on English language skills. The authors employ the similar methodology as in Angrist & Lavy (1997) with triple differences. In contrast to the Moroccan results by Angrist & Lavy (1997), they find no significant effect of the change of the language of instruction from English to Spanish on English skills of Puerto Ricans. One drawback of this study is that it does not focus on a particular labor market. Their data includes all people regarless of where they work, in the US or in Puerto Rico. Chakraborty & Bakshi (2012) make use of a law abolishing English courses in primary schools in some states of India. The authors use the variation across cohorts and districts to identify the effect of the law. Using a *two-way fixed effect* estimation, they find that those with less exposure to English courses earn significantly less even with the same level of education.

All three studies cited above estimate the effects of an environment with a dominant foreign language colonial language. In the case of Morocco it is French. In Puerto Rico and

India, it is English language. In a recent study, Anghel et al. (2012) look at the effect of the use of English-medium instruction in schools on students' academic performance in Spain. This paper is related to current paper in the sense that it deals with the effect of the language of instruction in an environment that English is not a dominant language in the labor market, and it is not a colonial language.

Anghel et al. (2012) make use of a special program adapted in primary schools in Spanish region of Madrid to look at the effect of English-medium instruction on learning. The educational authority allowed schools to give instructions in some subject in English language. The data set they use comes from a standardized exam taken by 6th graders. The data set has exam scores in three subjects "general knowledge," "reading" and "mathematics" for 2008/09 and 2009/10 school terms. Of three subjects, only general knowledge is taught in English. The authors use *difference-in-differences* approach to overcome the endogeneity problems. They consider two sources of endogeneity problems. The first one comes from the fact that the schools are not randomly selected to implement the bilingual program. The second source is the fact that students are not randomly assigned into schools. The paper finds that teaching in English has negative effect on General Knowledge for students whose parents have education level less than university, and no significant effects on Math and Reading which are taught in Spanish. One important caveat in their study is that the examination on General Knowledge, as on Math and Reading, is conducted in Spanish not in English. This may cause a reduction in General Knowledge scores of those who took the classes in English.

This study differs from the foregoing studies in important aspects. First of all, English is neither a colonial language in Turkey, nor a dominant language in the Turkish labor market. Our data set comes from a directed online survey that is designed and implemented for this research.[1] In that sense, the study makes use of a rich data set. In this paper, we focus on the monthly earnings of those who graduated from a four year undergraduate program, as opposed to primary or secondary educational institutions, with English-medium instruction. We came across no other study estimating the effect of the instructional language in university

[1] It "directed" in the sense that the survey link was not open to general public but sent only to the target population.

level. It is important in the sense that as the time lag between the graduation and the labor market entry gets longer, we may expect the effect of the language of instructions on labor market outcomes to wane.

3 Data and Some Descriptive Statistics

The data set used in this study comes from an online survey conducted between September 2017 and March 2018[2]. The target population is the people graduated between 2005 and 2017 from economics and business programs in two Turkish universities, Anadolu University and Dokuz Eylül University. Located in two different cities in Turkey, Anadolu University, Eskişehir and Dokuz Eylül University, İzmir, have both English-medium and Turkish-medium economics and business programs. Although the total number of graduates participated in the survey is 1325, approximately 13% of the target population, 227 of them have left important questions unanswered. We have dropped these missing observations. In the final sample of 1093 individuals, only 756 were employed at the time of survey, and 741 have provided earnings information.

Table 1: Programs and Institutions

Program	Anadolu Uni.	Dokuz Eylül Uni.	Whole sample
Economics (English)	26,27	87,84	44,72
Economics (Turkish)	42,39	7,60	31,97
Business (English)	7,80	3,95	6,65
Business (Turkish)	23,54	0,61	16,67
Total	100,00	100,00	100,00
	(769)	(329)	(1098)

Of 1098 graduates in the sample, 564 (51.4%) are from English-medium programs, and 534 (48.6%) are from Turkish-medium programs. The distribution of graduates among four programs and two universities are given in the Table 1. According to the table, There are 769 grad-

[2]Online survey is available at https://docs.google.com/forms/d/e/1FAIpQLSfK6g2So-qZtpdbHm-B_C0YKbwvNkVyXBVxteII3U7wVNcfSw/viewform.

uates in the sample from Anadolu University, and 329 graduates are from Dokuz Eylül University. The distribution of graduates in each universities and programs in the sample are as follows. Of those who graduated from Anadolu University, 26.3% are from English-medium economics, 42.4% are from Turkish-medium economics, 7.8% are from English-medium business and 23.5% are from Turkish-medium business programs. Of those who graduated from Dokuz Eylül University, 87.8% are from English-medium economics, 7.6% are from Turkish-medium economics, 3.9% are from English-medium business and 0.6% are from Turkish-medium business programs. 91.8% of Dokuz Eylül graduates in the sample are from English-medium programs. It is 34.1% among Anadolu University graduates.

Table 2: Summary Statistics

Variable	Obs.	Mean	Std. Dev	Min	Max
Earnings	741	3970,90	2835,00	360	28000
log(earnings)	741	8,11	0,57	5,89	10,23996
Male	1098	0,46	0,50	0	1
Age	1090	27,97	3,31	23	44
Married	1097	0,24	0,43	0	1
English program	1098	0,51	0,50	0	1
Anadolu Univ.	1093	0,70	0,46	0	1
Economics	1098	0,77	0,42	0	1
GPA	1078	2,73	0,43	2	3,94
Master's	1097	0,32	0,47	0	1
Empoloyed	1098	0,70	0,46	0	1
Experience	1091	3,60	2,89	0	14
Tenure (Year)	740	2,36	2,72	0	13
Internship	1098	0,51	0,50	0	1
Worked while being student	1098	0,35	0,48	0	1
Erasmus	1098	0,20	0,40	0	1
Computer	1098	0,28	0,45	0	1
Clubs	1098	0,50	0,50	0	1
Father's schooling (Year)	1095	10,45	4,07	0	18
Mother's schooling (Year)	1096	8,68	4,36	0	18
Father worked while studying	1096	0,92	0,27	0	1
Mother worked while studying	1096	0,40	0,49	0	1

Note: Region, the size of firm and the education category variables are not shown in this table.

In our sample, 756 respondents were employed at the time of the survey. Of employed respondents, 741 answered the question about labor earnings. In this regard, the maximum

Table 3: Graduates by high school types and programs

	Turkish-medium	English-medium	Whole sample
Anadolu/Soc. Science High	41,54	62,41	52,28
Anadolu Vocational	1,50	1,06	1,28
Vocational High School	3,76	1,42	2,55
Science High School	1,13	1,95	1,55
General High School	41,92	20,92	31,11
Foreign Language/Süper High	10,15	12,23	11,22
Total	100,00	100,00	100,00

number of observations that we can use in earnings equations is 741. The sample size has been further decreased due to the missing answers in various control variables. Nevertheless, the number of observations in all our analyses does not fall below 700. We have not included the variable "sector in which the mother works" in estimations since missing answers to this question significantly reduce the sample size. Table 2 provides the summary statistic of the variables used in our analyses.

Table 3 shows the distribution of graduates in the sample by their high schools and medium of instruction in undergraduate programs. With regard to English curriculum, Science High Schools Anadolu High Schools and Süper High Schools offer more intensive English classes compared to General and Vocational High Schools. Hence, graduates of these high schools are expected to have better English proficiency. The share of Vocational and Science High Schools are rather small in the sample. They consists of 5.38% of the whole sample. There is a noticeable difference between Anadolu High School and General High School backgrounds in English-medium programs. The share of Anadolu High School background in English-medium programs is three times higher than General High backgrounds. There appears no such difference in Turkish-medium programs comparing Anadolu High School and General High School backgrounds. The share of Anadolu High School and General High School bachgrounds in Turkish-medium programs are 41.5% and 42% respectively. The share of Anadolu High School backgrounds in Dokuz Eylül University programs is around 65% compared to 47% in Anadolu University.

The educational backgrounds of parents are given in Table 4. In both type of programs,

Table 4: Educational Status of Parents (%)

	Turkish-medium	English-medium	Whole sample
Father's Education			
Below high school	41,28	35,05	38,08
High school	29,83	27,76	28,77
University or above	28,89	37,19	33,15
Total	100,00	100,00	–
Mother's Education			
Below high school	61,35	45,12	53,01
High school	24,58	31,62	28,19
University or above	14,07	23,27	18,8
Total	100,00	100,00	–

mothers have lower level education compared to fathers, a reflecting the countrywise educational attainment levels by gender. The share of graduates with university educated fathers is around 37.2% in English-medium programs, and 29% in Turkish-medium programs. This pattern is also true for mothers. The numbers in Table 4 confirm that the same is true when we take educational attainment of parents as years of schooling. The parents of English medium graduates have higher levels of educational attainment than the parents of Turkish-medium graduates, and these differences are statistically significant.

Table 5 shows the differences of means in various variables between Turkish-medium and English-medium programs. The p values for the difference are in the last column. The overall result suggest that there exist statistically significant differences between Turkish-medium and English-medium graduates in the most of the characteristics in the table. Most notably, the average monthly earnings, both in levels and logs, are significantly higher for English-medium graduates.

4 Methodology

In order to see whether graduating from English medium undergraduate program has any effect on wages, we estimate a semi logarithmic wage equation. In the equation (1), w_i is the mothly earnings of individual i, X_i is a vector of various control variables, β is a vector of parameters, and ε_i is the random error term. D_i is a dummy variable indicating whether the

Table 5: Differences between Turkish and English-medium Programs

Variable	Turkish (mean)	English (mean)	Difference	p-value
Earnings	3261,40	4540,50	-1279,1***	0,000
log(earnings)	7,95	8,25	-0,296***	0,000
Male	0,47	0,46	0,00885	0,769
Age	27,80	28,10	-0,325	0,105
Married	0,23	0,25	-0,0191	0,460
Anadolu	0,95	0,46	0,488***	0,000
Dokuz Eylül	0,05	0,54	-0,488***	0,000
Economics	0,66	0,87	-0,213***	0,000
GPA	2,67	2,78	-0,106***	0,000
Master's	0,29	0,36	-0,0687*	0,015
Unemployed	0,36	0,24	0,120***	0,000
Employed	0,64	0,76	-0,120***	0,000
Experience	3,48	3,71	-0,226	0,198
Tenure (Year)	2,44	2,29	0,152	0,452
Public Sector	0.17	0.15	0.0250	0.256
Self employed	0.02	0.02	0.00297	0.732
Private Sector	0.43	0.58	-0.147***	0.000
Internship	0,41	0,62	-0,211***	0,000
Worked while student	0,38	0,32	0,0683*	0,018
Erasmus	0,09	0,31	-0,219***	0,000
Computer	0,31	0,24	0,0681*	0,012
Clubs	0,48	0,51	-0,0294	0,331
Father's schooling (Year)	10,1	10,80	-0,676**	0,006
Mother's schooling (Year)	8,07	9,25	-1,179***	0,000
Father worked while student	0,91	0,92	-0,0102	0,536
Mother worked while student	0,38	0,42	-0,0347	0,242
Father's Occupation				
High skill/white collar	0,33	0,53	-0,201***	0,000
Low skill/white collar	0,25	0,18	0,0742**	0,003
High skill/blue collar	0,02	0,004	0,0154*	0,015
Low skill/blue collar	0,01	0,007	0,00236	0,667
Unemployed/missing	0,39	0,28	0,109***	0,000
Mother's Occupation				
High skill/white collar	0,33	0,40	-0,0658*	0,028
Low skill/white collar	0,14	0,10	0,0413*	0,042
High skill/blue collar	0,23	0,20	0,0296	0,243
Low skill/blue collar	0,13	0,17	-0,0356	0,109
Unemployed/missing	0,17	0,14	0,0304	0,177

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

individual was graduated from an English medium program. Our parameter of interest is τ which can be estimated using OLS.

$$\log\, w_i = \alpha + \tau D_i + X_i'\beta + \varepsilon_i \tag{1}$$

$$D_i = \begin{cases} 1 & : \text{Graduated from English-medium program} \\ 0 & : \text{other} \end{cases}$$

In estimating a model such as (1), there are two possible sources of bias. The first is due to *sample selection*. The sample selection problem arises from the fact that we can observe earnings only for the employed respondents. If employed and unemployed individuals systematically differ in some unobserved characteristics such as abilityi motivation or professional commitment that might affect the reservation wage of individuals, estimated parameters from OLS will be biased. Sample selection one of the problems faced in econometric studies aiming to establish a causal relationship (Manski, 1995, pp. 21–50). We try to remedy any possible sample selection bias by estimating a two equation *"Heckman Selection"* model (Gronau, 1973; Heckman, 1974; Maddala, 1983; Mroz, 1987) :

$$\log\, w_i = \alpha + \tau D_i + X_i'\beta + \varepsilon_i \qquad \text{(for } s_i = 1\text{)} \tag{2}$$

$$s^* = Z_i'\gamma + u_i \quad s_i = 1[s^* > 1] \tag{3}$$

$$(u_i, \varepsilon_i) \sim N\,(0, 0, \sigma_u, \sigma_\varepsilon, \sigma_{\mathrm{u}}\,) \qquad \text{and} \qquad E(u|Z) = E(\varepsilon|X) = 0$$

In model above, equation (3) is the *"selection equation"*, and equation (2) is the *"output equation"* (Greene, 2011, p. 875). s_i^* is a latent variable, while s_i is an indicator variable that takes 1 for the employed. We assume that error terms u_i and ε_i have a bivariate normal distribution. One can think of the latent variable as the difference between the reservation wage of the individual and the offered wage in the labor market. A positive difference signals that an individual searching for a job will accept the offer and get employed. The monthly earning variable is only observed for those who are employed, $s_i = 1$. This model can be estimated with *full information maximum likelihood* (FIML) or *limited information maximum likelihood* (LIML) method, which is also known as the "two-stage" method. The two-stage method uses the *inverse Mill's ratio* obtained from estimating the selection equation as an explanatory variable in the output equation.

In the estimation of the model, the vector Z usually includes at least one additional variable that is not included in vector X. This is called **exclusion restrictions**. For identification of the parameters in the FIML estimation, theoretically there is no need for an exclusion restriction and X can be equal to Z. In this case, identification relies on bivariate normality assumption. However, in case of a highly linear inverse Mill's Ratio, multicollinearity problems may arise. Hence, exclusion restrictions should be used where possible.

The second possible source of bias is related to the *"endogenous selection"* which arises as a result of endogenous sorting by some observed or unobserved factors instead of randomly assigning students into programs. This problem of selection will cause identification problem, and estimated parameters will be biased. Endogenous selection may be caused by some unobserved variables such as ability and communication skills, as well as some observed variables such as socio-economic status, gender, region. In endogenous selection, unlike sample selection, the dependent variable of the outcome equation, monthly earnings in our case, is observed for all individuals.

We adopt the *Potential Outcome Model* (POM) to describe endogenous selection problem in our context[3]. Consider the where some high school graduates choose to major an English medium economics program, while some others choose the Turkish medium economics program of the same university. Students who graduted from Turkish medium program consist of the control group, while students graduated from the English medium program are the treatment group. Let y_1 be the future earnings of students that graduated from the English medium program, while y_0 are the future earnings of those who graduated from the Turkish

[3] *Potential Output Model* or *Rubin's Causality Model* approach, first developed by Rubin (1974), is widely used in analyzing causal relationahips (Angrist, Imbens, & Rubin, 1996; Heckman, 2010; Heckman & Vytlacil, 2005; Imbens & Wooldridge, 2009)

medium program. Let D be a binary *treatment* or *participation* variable taking value 1 if student is graduated from English medium program and value 0 otherwise. For student i the *Treatment Effect* (TE) refers to the difference between the earning of students participating in and those not participating in the program:[4]

$$TE_i = y_{i1} - y_{i0}$$

TE indicates the effect of treatment on individual i at only one time. Obviously it is not possible to simultaneously observe both results y_1 and y_0 for the same individual. The students could be either in the Turkish medium program, and thus have an observed outcome y_0, or in the English medium program, with observed outcome y_1.[5] Therefore, we can only observe the *"potential output"*:

$$y_i = (1 - D_i)y_{i0} + D_i y_{i1} = y_{i0} + D_i(y_{i1} - y_{i0})$$

The average treatment effect (ATE) is defined as follows:

$$\tau_{\text{ATE}} \equiv E(y_1 - y_0) \tag{4}$$

ATE expresses the average (expected) effect of the earnings of a randomly selected student from the English medium program. The treatment effect on only on those students who, in fact, are enrolled in the program is called *the average treatment on treated* (ATT):

$$\tau_{\text{ATT}} \equiv E(y_1|D = 1) - E(y_0|D = 1) = E(y_1 - y_0|D = 1)$$

[4] Here, we follow the notation by Wooldridge (2010).

[5] We exclude the case of double major in two programs that are included in the data set.

Since ATE is the observed average treatment effect, equation (4) can be written as follows:

$$\begin{aligned}
\tau_{\text{ATE}} &\equiv E(y_1|D=1) - E(y_0|D=0) \\
&= E(y_1|D=1) - E(y_0|D=0) - E(y_0|D=1) + E(y_0|D=1) \\
&= E(y_1|D=1) - E(y_0|D=1) + E(y_0|D=1) - E(y_0|D=0) \\
&= \tau_{\text{ATT}} + E(y_0|D=1) - E(y_0|D=0)
\end{aligned} \tag{5}$$

In equation (5), if the assignment variable D_i is not independent of the earnings variable, i.e $E(y_0|D=1) \neq E(y_0|D=0)$, then the ATE estimates will be biased. If, on the other hand, the assignment variable D and the earnings (y_1, y_0) are independent (i.e in the case of randomized enrollment), then τ_{ATE} and τ_{ATT} will be equal and ATE will simply show the difference between the means of the treatment and the control groups:

$$\begin{aligned}
&E(y|D=1) = E(y_1|D=1) = E(y_1) \\
&E(y|D=0) = E(y_0|D=0) = E(y_0) \\
&\qquad \tau_{\text{ATT}} \equiv E(y_1 - y_0|D=1) = E(y_1 - y_0) \equiv \tau_{\text{ATE}}
\end{aligned}$$

The problem of endogenous selection may arise in two ways. The first of these is called *"observable selection"* where the variable (confounder) that establish the relationship between the assignment variable and the earnings variable is observable. In this case, including the confounder in the regression equation results in unbiased and consistent OLS estimates. The observable selection depends on two assumptions: *the assumption of conditional independence* and *the assumption of overlap* (Wooldridge, 2010, Ch. 21). Under the conditional independence, it is assumed that the confounding variables that determine the selection are X variables, and when controlled, the assignment variable and the earnings variable are independent:

$$(y_1, y_0) \perp D | X$$

The conditional independence is a very strong assumption. In practice, usually a weaker version, the average conditional independence is assumed:

$$E(y_1|X, D) = E(y_1|X)$$

$$E(y_0|X, D) = E(y_0|X)$$

The overlap assumption implies that the probability of an individual graduating from the English medium program in the sample is not 0 or 1:

$$0 < P(D_i = 1|X_i) < 1$$

If the conditional independence assumption is violated, then there must exist an unobserved confounding variable that affects both assignment and the outcome . In this case, we have an *unobservable selection* problem. There is no formal test for the violation of conditional independence. Therefore, in both cases, we need to follow economic arguments.

In case of observable seleciton, Matching, Regression Correction (RA), Inverse Probability Weighting (IPW) estimators are possible alternatives. All are based on the conditional independence and overlap assumptions. If these assumptions are violated, estimations will be biased.In case of unobservable selection, the first method that can be used is instrumental variable (IV). However, it is not always possible to find a valid instrument. In this case, the methods that do not need the *exclusion restrictions* are available. These include bivariate normal selection model, control function method and minimum deviation method. In this study, we estimate only the bivariate normal selection model.

4.0.1 Matching Methods

Matching methods are commonly used estimation methods in the case of observed selections[6]. The idea of matching methods is to find identical or very similar observations from the treatment and the control group, and determine the extent to which they differ in terms of treatment results. The process of identifying similar observations is called "matching". With observed data the probability of having exactly identical observations in both treatment and control groups is very low. In practice, matching only similar observations is possible. We employ two methods here, the *nearest neighbor* (NN), and the *propensity score* (PS). In nearest neighbor matching, the observations in a close distance with respect to control variables are compared (Abadie & Imbens, 2006, 2011; Wooldridge, 2010).[7] Various metrics are available to determine the distance measure. We use *Mahalanobis norm* in NN estimations here (see Rubin, 1979; Gu & Rosenbaum, 1993; Wooldridge, 2010, p. 934).

Propensity score estimator matches observations with similar propensity scores (Abadie & Imbens, 2016; Rosenbaum & Rubin, 1983; Wooldridge, 2010, p. ch. 21). The propensity score p is simply the probability of an individual i being treated:

$$P(D = 1|X) = p(X)$$

Logit or probit models can be used to obtain the propensity scores. Propensity score matching gives better results when the number of conrtol variables is large (Gu & Rosenbaum, 1993). If some control variables are continuous, then we need an additional assupmtion, *balancing*, be satisfied. The balancing assumption requires that the distribution of control variables is balanced between the treatment and the control groups after matching propensity scores (Cerulli, 2015, p. 70; Gu & Rosenbaum, 1993):

[6]See matching literature and its appliation (see Caliendo & Kopeinig (2008); Stuart (2010)

[7]Wooldridge (2010) refers to this matching as "matching on covariates" or "covariate matching".

$(D \perp X)|p(X)$

4.0.2 Regression Adjustment

The regression adjustment (RA) is the simplest method that can be used in the case of observed selection. The RA method estimates the output equation separately for the treatment and the control group, and takes their average. The method gives an average estimation of ATE. Let us express the conditional expected values of the output variable with the equations $m_1(X) = E(y|X, D = 1)$ and $m_0(X) = E(y|X, D = 0)$ for the treatment and the control group respectively. If $\widehat{m}_1$ and $\widehat{m}_0$ are the consistent estimates of these expected values from a random sample with size N, then the regression adjustment for ATE can be expressed as:

$$\widehat{\tau}_{\text{ATE}} = N^{-1} \sum_{i=1}^{N} [\widehat{m}_1(X) - \widehat{m}_0(X)]$$

$\widehat{m}_1$ and $\widehat{m}_0$ can be estimated using parametric or non-parametric methods. We use the predicted values from OLS by assuming a linear relationship (Imbens & Wooldridge, 2009; Wooldridge, 2010, ch. 21). Busso, DiNardo, & McCrary (2014) show that the regression adjustment method performs better in samples with high overlapping than matching methods.

4.0.3 Inverse Probability Weighting

Another estimation method used to remedy observed selection bias is the inverse probability weighting (IPW). IPW gives a lower weight to the observations that are more likely to be treated and a higher weight for those who are less likely to be treated. If we express the probability of being English medium graduate by $P(D = 1|X_i)| \equiv P(X_i)$, then the probability of being Turkish medium graduate will be $1 - P(X_i)$. The IPW estimation of ATE is given by the following (Wooldridge, 2010, p. 823):

$$\hat{\tau}_{\text{ATE}} = \frac{\sum_{i=1}^{N} \frac{D_i y_i}{\hat{p}(X_i)}}{\sum_{i=1}^{N} \frac{D_i}{\hat{p}(X_i)}} - \frac{\sum_{i=1}^{N} \frac{(1-D_i) y_i}{1-\hat{p}(X_i)}}{\sum_{i=1}^{N} \frac{D_i}{1-\hat{p}(X_i)}} \tag{6}$$

In equation (6), $\widehat{p}(X_i)$ is the propensity score obtained from the sample, N is the sample size, y is the monthly earnings, D is the treatment variable and X is the vector of control variables.

4.0.4 Endogenous Treatment Model

The selection model proposed by Heckman (1978, 1979) which is used to eliminate the sample selection bias, can also be used in the case of endogenous selection without truncation. The difference in this case is that we observe the outcome variable for both selected and not selected individuals. Identification in endogenous treatment model heavily depends on the bivariate normal distribution assumption about the error terms in outcome and selection equations:

$$\log w_i = \alpha + \tau D_i + X_i' \beta + \varepsilon_i \tag{7}$$

$$D_i = 1[Z_i' \gamma + u_i > 0] \tag{8}$$

$$(u_i, \varepsilon_i) \sim N\left(0, 0, \sigma_u, \sigma_\varepsilon, \sigma_{\text{u}}\right) \qquad \text{and} \qquad E(u|Z) = E(\varepsilon|X) = 0$$

In equation (8), D is the binary treatment variable taking 1 if individual i is an English medium graduate. This is the model 5 in Maddala (1983, p. 120). We model the selection equation as a probit, estimate the model using FIML. The identification in this model depends entirely on the bivariate normality assumption. Hence, no exclusion restrictions are required, and $Z = X$.

5 Estimation Results

Our base estimates obtained using the OLS method are presented in Table 7. We estimated four different models, numbered (1)-(4) in the table. All models include regional dummies and constant terms. Model (1) does not include parents' education and variables such as sector, occupation and firm size. Angrist and Pischke (2014, p. 217) recommend against to control variables that are determined by the teatment variable. This is due to the fact that these are outcome variables determined by the treatment variable rather than control variables. Including these variables in the model may impose endogeneity into the model. Angrist & Pischke (2014) refers to these variables as *bad controls*. In our context, bad control are occupations, sectors, firm size, and extracurricular activities during university years, such as participating in Erasmus exchange programs, internships, and participating student club events. Since these variables have a positive effect on earnings, including them in the model will lead to a decrease in the size of average treatment effect τ.

Adding these variables to the model might be useful to see a lower bound for the ATE estimations. In fact, the estimated ATE in Model (1) is 21%, while in (3) and (4), it falls down to 11%. The education of parents does not appear to have a significant impact on the ATE estimates. Therefore, when potential selection bias is ignored, we can say that graduating from an English-medium program leads to an increase in the average monthly earnings by at least 11%. Looking at other variables affecting monthly earnings, such as being male, married, economics major and Dokuz Eylül University graduate significantly increase average monthly earnings. The estimated parameters for experience and experience-squared show that earnings rise with more experience, while the marginal contribution of this rise is decreasing. The estimation results show that internship has no significant effect on earnings. In both of the universities, the internship is optional, and a matter of personal choice. Traveling and studying abroad via Erasmus programs, on the other hand, has a positive effect on earnings.

Another point that should be noted is related to the sector of employment in OLS estimates. We set the public sector as a reference. After controlling other factors, self-employment does not have any significant effect on earnings compared to the public sector. Furthermore, graduates working in the private sector have lower monthly earnings compared to those working in the public sector.

In order to control for a possible sample selection, we estimated Heckman correction model using FIML. The estimations from the outcome equations are in Table 8, and the results from the selection equation are in Table 9. The selection equations are probit models with the dependent variable being employment dummy. Statistically significant *atrho* parameter in all models indicates that the outcome and the selection equations are not independent.[8] This implies that, under the assumption of a bivariate normal distribution of the error terms, there exists a sample selection bias. In other words, the sample selection models are, in general, significant. On the other hand, it is not possible to test whether the models meet the bivariate normal distribution assumption which is the basic assumption of these models.The ATE estimates are given by the coefficient of English-medium dummy variable. The estimated coefficient of this variable is somewhat lower than the corresponding OLS estimates. According to the model without bad controls and parents's education, having graduated from an English-medium program raise monthly earnings by 18%. This is the highest estimate in all four models given by Table 8. Model (4) in this table produces the smallest estimation, 10%, for ATE of English-medium programs. The standart errors in sample selection models are higher than OLS models as expected.

Table 10 presents the outcome equations from Heckman type endogeneous treatment models. We estimated all four models using FIML. Note that in these models, the dependent variables in the selection equations is English-medium dummy. $atan\rho$ estimate is not significant in all models except Model (3). Moreover, the standart errors of estimated coeffiencts are high, so English-medium coefficient is not stiatistically significant. Although the selection equations are significant as a whole, many parameters appears not different from zero (Table 11). The exclusion restrictions in all models include high school type, whether mother was working while undergraduate education, the reagion lived during high school and whether the respondent had English language skills before undergraduate education. Hence, the models

[8] The "*athro*" ($atan\rho$) parameter in the table is defined as $\mathrm{atan}\rho = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$.

are overidentified.

In addition to sample selection in analysis, we estimate various models to control for observed and unobserved internal selection. We performed Regression Adjustment (RA), Inverse Probability Weighting (IPW), Propensity Score Matching (PSM) and Nearest Neighbor Matching (NNM) estimations for possible observable selection bias. The ATE and ATT estimates from observed selection models are given in Table 6. In the second and third columns of the table are the ATE and ATT estimates respectively. The first thing that draws attention in these estimates is that all observable selection models give greater coefficient estimates than OLS estimates. Moreover, all observed selection models yielded very close ATE estimates. Figure 1 shows the balance of propensity score estimates. The control variables are distributed in an even way between the treatment (English-medium) and control (Turkish-Medium) groups after propensity score matching. In other words, propensity score matching choses similar observations for the treatment and the control groups.

If there exists a selection in the sample used in the analysis and the assumption of conditional independence is satisfied, the estimations obtained from the RA, IPW, PSM and NNM models will be unbiased and consistent. It is not possible to test the assumption of conditional independence. Therefore, it is difficult to determine which model produces consistent results. However, considering very close estimated ATE coefficient yielded by different models, we can argue that the selection may not be an important sources of bias. This is plausable for a couple of reasons. First of all, the existing sample is drawn from a narrow demographic group. All individuals in the sample are graduates of two universities in cities with similar demographics, and have on average 3.6 years of labor market experience (Table 2). The mean age of the sample is 27 and the standard deviation of age is 3.3 years. Therefore, this is a sample with somewhat similar demographic characteristics. Another important point is that English-medium and Turkish-medium programs in Anadolu University are offered by the same department and tought by the same faculty. Thus, many unobservable factors about university, city and demographics are already controlled during the sampling process.

6 Conclusion

The effect of graduating from an English-medium undergraduate program on earnings in the labor market is an important matter of discussions on the opening of universities that provide education in English. If English-medium programs does not have any positive effect on labor market ourcomes, then these programs will only serve for other purposes such as a competition tool in attracting high ability students, or socioeconomic sorting. Therefore, finding answers to this question is of great importance. This is the first study to examine the effect of English-medium instruction in higher education on labor market earnings in an economy where English is not widely spoken or a colonial language. The data set used in this study allows us to properly estimate this effect.

Endogeneity is a major problem in any analysis of the effects of various factors on earnings. Endogeneity leads to identification problem and yields biased estimates. In this study, we use different econometric tools to take into account the endogenity resulting from sample selection, observable seleciton and unobservable selection. All selection models estimated yielded higher ATE estimates compared to OLS estimates. It is difficult to predict the direction and size of possible selection bias in the model. The main reason for this is that there may exist more than one source of endogeneity at the same time. Therefore, it is difficult to interpret differences between the OLS and other estimates. However, in the case of a selection, OLS estimates can be interpreted as a lower bound of ATE. In this regard, we find that the average monthly earnings difference between the graduates from English-medium and Turkish-medium programs vary between 11% and 22%.

References

Abadie, A., & Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, *74*(1), 235–267. https://doi.org/10.1111/j.1468-0262.2006.00655.x

Abadie, A., & Imbens, G. W. (2011). Bias-corrected matching estimators for average treatment effects. *Journal of Business & Economic Statistics*, *29*(1), 1–11. https://doi.org/10.1198/jbes.2009.07333

Abadie, A., & Imbens, G. W. (2016). Matching on the estimated propensity score. *Econometrica*, *84*(2), 781–807. https://doi.org/10.3982/ecta11293

Anghel, B., Cabrales, A., & Carro, J. M. (2012). *Evaluating a bilingual education program in spain: The impact beyond foreign language learning* (Economics Working Papers No. we1214). Universidad Carlos III, Departamento de Economía. Retrieved from http://ideas.repec.org/p/cte/werepe/we1214.html

Angrist, J., Chin, A., & Godoy, R. (2008). Is spanish-only schooling responsible for the puerto rican language gap? *Journal of Development Economics*, *85*(1-2), 105–128. Retrieved from http://ideas.repec.org/a/eee/deveco/v85y2008i1-2p105-128.html

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444–455. https://doi.org/10.1080/01621459.1996.10476902

Angrist, J. D., & 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), S48–76. Retrieved from http://ideas.repec.org/a/ucp/jlabec/v15y1997i1ps48-76.html

Angrist, J. D., & Pischke, J.-S. (2014). *Mastering metrics: The path from cause to effect*. Princeton Univers. Press. Retrieved from https://www.ebook.de/de/product/22417805/angrist_jorn_steffen_pischke_mastering_metrics_the_path_from_cause_to_effect.

html

Bleakley, H., & Chin, A. (2004). Language skills and earnings: Evidence from childhood immigrants. *The Review of Economics and Statistics*, *86*(2), 481–496. Retrieved from http://ideas.repec.org/a/tpr/restat/v86y2004i2p481-496.html

Busso, M., DiNardo, J., & McCrary, J. (2014). New evidence on the finite sample properties of propensity score reweighting and matching estimators. *Review of Economics and Statistics*, *96*(5), 885–897. https://doi.org/10.1162/rest_a_00431

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*(1), 31–72. https://doi.org/10.1111/j.1467-6419.2007.00527.x

Cerulli, G. (2015). *Econometric evaluation of socio-economic programs*. Berlin: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-46405-2

Chakraborty, T., & Bakshi, S. K. (2012). *English language premium: Evidence from a policy experiment in india*. Unpublished Manuscript.

Coleman, J. A. (2006). English-medium teaching in european higher education. *Language Teaching*, *39*(1), 1–14. https://doi.org/10.1017/s026144480600320x

Ekoç, A. (2018). English medium instruction (EMI) from the perspectives of students at a technical university in turkey. *Journal of Further and Higher Education*, 1–13. https://doi.org/10.1080/0309877x.2018.1527025

Greene, W. H. (2011). *Econometric analysis* (7th ed.). Boston: Pearson.

Gronau, R. (1973). The effect of children on the housewifes value of time. *Journal of Political Economy*, *81*(2, Part 2), S168–S199. https://doi.org/10.1086/260160

Gu, X. S., & Rosenbaum, P. R. (1993). Comparison of multivariate matching methods: Structures, distances, and algorithms. *Journal of Computational and Graphical Statistics*, *2*(4), 405–420. https://doi.org/10.2307/1390693

Heckman, J. (1974). Shadow prices, market wages, and labor supply. *Econometrica*, *42*(4), 679–694. https://doi.org/10.2307/1913937

Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica*, *46*(4), 931–959. https://doi.org/10.2307/1909757

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153. https://doi.org/10.2307/1912352

Heckman, J. J. (2010). Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic Literature*, *48*(2), 356–398. https://doi.org/10.1257/jel.48.2.356

Heckman, J. J., & Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, *73*(3), 669–738. https://doi.org/10.1111/j.1468-0262.2005.00594.x

Hu, G., & Alsagoff, L. (2010). A public policy perspective on english medium instruction in china. *Journal of Multilingual and Multicultural Development*, *31*(4), 365–382. https://doi.org/10.1080/01434632.2010.489950

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*(1), 5–86. https://doi.org/10.1257/jel.47.1.5

Kılıçkaya, F. (2006). Instructors' attitudes towards english-medium instruction in turkey. *Humanising Language Teaching*, *8*(6), 1–8. Retrieved from http://files.eric.ed.gov/fulltext/ED570169.pdf

Kırkgöz, Y. (2014). Students' perceptions of english language versus turkish language used as the medium of instruction in higher education in turkey. *Turkish Studies*, *9*(12), 443–459.

Lang, K., & Siniver, E. (2006). *The return to english in a non-english speaking country: Russian immigrants and native israelis in israel* (Working Paper No. 12464). National

Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w12464

Levinsohn, J. (2004). *Globalization and the returns to speaking english in south africa* (NBER Working Papers No. 10985). National Bureau of Economic Research, Inc. Retrieved from http://ideas.repec.org/p/nbr/nberwo/10985.html

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9780511810176

Manski, C. F. (1995). *Identification problems in the social sciences*. Cambridge: Harvard University Press.

Mroz, T. A. (1987). The sensitivity of an empirical model of married womens hours of work to economic and statistical assumptions. *Econometrica*, *55*(4), 765. https://doi.org/10.2307/1911029

Munshi, K., & 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.

Özen, E. N., Alpaslan, I. B., Çağlı, A., Özdoğan, I., Sancak, M., Dizman, A. O., & Sökmen, A. (2014). *Turkey national needs assessment of state school english language teaching*. Report, Ankara: TEPAV; British Council; British Council. Retrieved from https://www.tepav.org.tr/upload/files/1399388356-5.Turkey_National_Needs_Assessment_of_State_School_English_Language_Teaching.pdf

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701. https://doi.org/10.1037/h0037350

Rubin, D. B. (1979). Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association*, *74*(366), 318. https://doi.org/10.2307/2286330

Selvi, A. F. (2014). The medium-of-instruction debate in turkey: Oscillating between national ideas and bilingual ideals. *Current Issues in Language Planning*, *15*(2), 133–152. https://doi.org/10.1080/14664208.2014.898357

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, *25*(1), 1–21. https://doi.org/10.1214/09-sts313

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). Cambridge: The MIT Press.

Appendix

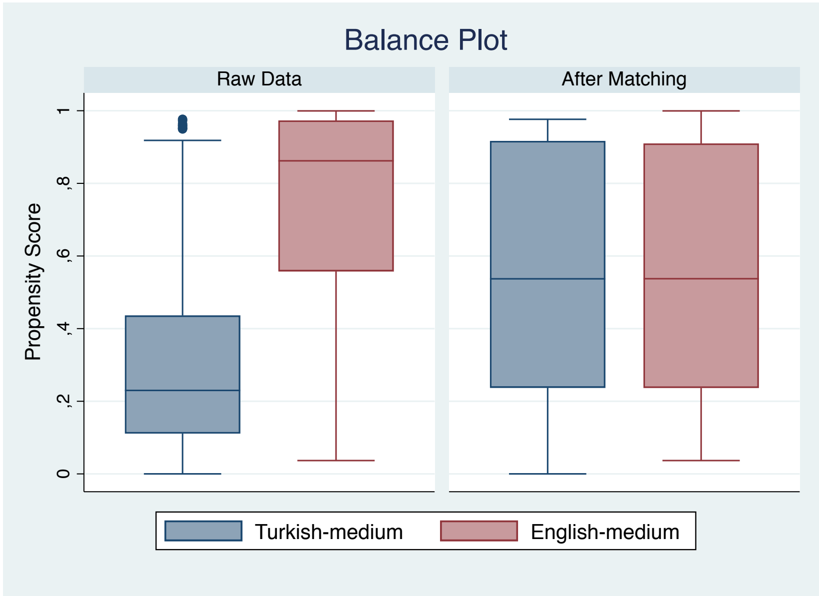

Figure 1: Matching Balance

Table 6: Observable Selection Model Estimates

Model	ATE	ATT
Regression Adjustment	0,196***	0,202**
Inverse Probability Weighting	0,188***	0,210**
Propensity Score Matching	0,225***	0,313***
Nearest Neighbor Matching	0,205***	0,218***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: OLS Estimation Results

	(1)	(2)	(3)	(4)
English-medium	0,212***	0,210***	0,115***	0,117***
	(6,09)	(5,90)	(3,33)	(3,33)
Male	0,216***	0,219***	0,192***	0,193***
	(6,88)	(6,94)	(6,18)	(6,24)
Married	0,0851*	0,0855*	0,100*	0,0991*
	(2,06)	(2,06)	(2,58)	(2,54)
Economics	0,112**	0,118**	0,0903*	0,0943*
	(2,75)	(2,94)	(2,33)	(2,44)
Anadolu	-0,100*	-0,0956*	-0,0985*	-0,0938*
	(-2,31)	(-2,20)	(-2,40)	(-2,30)
Experience	0,141***	0,140***	0,122***	0,121***
	(7,44)	(7,46)	(6,62)	(6,61)
Experience^2	-0,00629***	-0,00632***	-0,00512**	-0,00507**
	(-3,98)	(-4,05)	(-3,30)	(-3,29)
Full time	0,544***	0,543***	0,427***	0,421***
	(4,46)	(4,58)	(3,69)	(3,73)
Self-employed			0,249	0,251
			(1,81)	(1,74)
Private sector			-0,107**	-0,105**
			(-2,68)	(-2,59)
Master's			0,0812*	0,0773*
			(2,23)	(2,10)
Erasmus			0,117**	0,109**
			(3,14)	(2,89)
Internship			0,0271	0,0227
			(0,84)	(0,70)
Firm size (categoric)	–	–	✓	✓
Occupation (categoric)	–	–	✓	✓
Father Education (categoric)	–	✓	–	✓
Mother Education (categoric)	–	✓	–	✓
N	721	719	715	713
R-squared	0,494	0,500	0,570	0,574

Notes: t statistics in parentheses. All models include constant term and regional dummies.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: Sample Selection Model Outcome Equations

	(1)	(2)	(3)	(4)
Log earnings				
English-medium	0.180***	0.179***	0.100**	0.102**
	(4.73)	(4.68)	(2.75)	(2.79)
Male	0.186***	0.186***	0.168***	0.168***
	(5.75)	(5.75)	(5.35)	(5.37)
Married	0.0451	0.0469	0.0622	0.0622
	(1.05)	(1.11)	(1.54)	(1.55)
Economics	0.0882*	0.0950*	0.0729	0.0771
	(2.13)	(2.32)	(1.81)	(1.94)
Anadolu Univ.	-0.0919*	-0.0881*	-0.0901*	-0.0862*
	(-2.14)	(-2.05)	(-2.25)	(-2.17)
Experience	0.0994***	0.0991***	0.0860***	0.0858***
	(4.43)	(4.54)	(4.01)	(4.11)
Experience^2	-0.00550**	-0.00559***	-0.00474**	-0.00478**
	(-3.18)	(-3.32)	(-2.84)	(-2.92)
Tenure	0.00253**	0.00254**	0.00265**	0.00263**
	(2.77)	(2.88)	(2.94)	(2.99)
Full time	0.472***	0.474***	0.351**	0.348**
	(3.92)	(4.05)	(3.15)	(3.22)
Self employed			0.289*	0.283*
			(2.19)	(2.11)
Private sector			-0.0910*	-0.0866*
			(-2.26)	(-2.13)
Master's			0.0916*	0.0884*
			(2.36)	(2.26)
Erasmus			0.0720	0.0612
			(1.81)	(1.51)
Internship			0.00546	0.000165
			(0.17)	(0.01)
GPA			0.0380	0.0434
			(0.90)	(1.03)
Firm size (categoric)	–	–	✓	✓
Occupation (categoric)	–	–	✓	✓
Father Education (categoric)	–	✓	–	✓
Mother Education (categoric)	–	✓	–	✓
athrho	-0.714***	-0.719***	-0.622*	-0.639**
	(-3.41)	(-3.68)	(-2.44)	(-2.77)
$\ln\sigma$	-0.852***	-0.858***	-0.954***	-0.955***
	(-14.14)	(-14.55)	(-14.10)	(-14.73)
Observations	1037	1037	1032	1032

Notes: t statistics in parentheses. All models include constant term and regional dummies.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9: Sample Selection Model Selection Equation

	(1)	(2)	(3)	(4)
Employed				
Male	0.278**	0.279**	0.267**	0.268**
	(2.90)	(2.93)	(2.76)	(2.79)
Married	0.352**	0.352**	0.335*	0.335*
	(2.63)	(2.63)	(2.50)	(2.50)
English-medium	-0.00530	-0.000697	0.0479	0.0478
	(-0.05)	(-0.01)	(0.48)	(0.48)
Master's	-0.268*	-0.264*	-0.383***	-0.376***
	(-2.31)	(-2.29)	(-3.72)	(-3.65)
Age	0.398	0.398	0.439	0.431
	(1.69)	(1.71)	(1.90)	(1.90)
Age^2	-0.00414	-0.00412	-0.00482	-0.00469
	(-1.02)	(-1.03)	(-1.21)	(-1.19)
Internship	0.211*	0.211*	0.215*	0.217*
	(2.25)	(2.26)	(2.20)	(2.23)
Erasmus	0.475***	0.472***	0.373**	0.381**
	(3.87)	(3.87)	(3.04)	(3.10)
GPA	0.225*	0.228*	0.155	0.154
	(2.06)	(2.10)	(1.38)	(1.37)
High school				
Anadolu Voc.	0.131	0.132	-0.0265	-0.0372
	(0.33)	(0.33)	(-0.06)	(-0.09)
Vocational	0.0372	0.0273	0.111	0.106
	(0.14)	(0.10)	(0.40)	(0.38)
Science high	-0.271	-0.269	-0.331	-0.332
	(-0.68)	(-0.68)	(-0.87)	(-0.87)
General high	0.0214	0.0148	-0.000511	-0.00256
	(0.19)	(0.13)	(-0.00)	(-0.02)
Foreign lan/Sup.	-0.0429	-0.0475	-0.0187	-0.0191
	(-0.26)	(-0.29)	(-0.11)	(-0.11)
Previous English	0.244*	0.243*	0.215*	0.216*
	(2.30)	(2.28)	(1.96)	(1.97)
Mother worked	-0.0170	-0.0155	-0.0231	-0.0217
	(-0.16)	(-0.15)	(-0.21)	(-0.20)
Father Education (categoric)	–	✓	–	✓
Mother Education (categoric)	–	✓	–	✓
LR χ^2	229,43	229,43	228,94	228,94
p-value	0,000	0,000	0,000	0,000
Observations	1037	1037	1032	1032

Notes: t statistics in parentheses. All models include constant term and regional (at high school) dummies.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: Heckman Endogenous Treatment Model Outcome Equation

	(1)	(2)	(3)	(4)
Log earnings				
English-medium	0.216	0.214	0.120	0.121
	(1.80)	(1.45)	(0.97)	(0.74)
Male	0.211***	0.215***	0.193***	0.197***
	(6.92)	(7.00)	(6.50)	(6.55)
Married	0.0711	0.0726	0.0916*	0.0916*
	(1.84)	(1.88)	(2.50)	(2.49)
Economics	0.113**	0.119**	0.0930*	0.0968*
	(2.83)	(2.97)	(2.43)	(2.52)
Anadolu Univ.	-0.0933*	-0.0864*	-0.0893*	-0.0823*
	(-2.19)	(-2.02)	(-2.21)	(-2.02)
Experience	0.124***	0.123***	0.105***	0.105***
	(6.65)	(6.56)	(5.91)	(5.83)
Experience^2	-0.00641***	-0.00651***	-0.00539***	-0.00548***
	(-4.48)	(-4.50)	(-3.94)	(-3.95)
Tenure	0.00248***	0.00255***	0.00265***	0.00267***
	(3.36)	(3.39)	(3.75)	(3.71)
Full time	0.494***	0.496***	0.354***	0.350***
	(5.77)	(5.79)	(4.15)	(4.09)
Sector				
Self employed			0.297**	0.286**
			(3.13)	(2.99)
Private sector			-0.0892*	-0.0863*
			(-2.20)	(-2.12)
Mater's			0.0650	0.0595
			(1.89)	(1.72)
Erasmus			0.103*	0.0934
			(1.99)	(1.53)
Intership			0.0262	0.0191
			(0.72)	(0.49)
GPA			0.0461	0.0530
			(1.24)	(1.40)
Firm size (categoric)	–	–	✓	✓
Occupation (categoric)	–	–	✓	✓
Father Education (categoric)	–	✓	–	✓
Mother Education (categoric)	–	✓	–	✓
athrho	-0.0936	-0.696***	-0.0289	-0.0276
	(-0.60)	(-6.06)	(-0.12)	(-0.09)
$\ln \sigma$	-0.918***	-0.858***	-0.994***	-0.994***
	(-34.10)	(-22.76)	(-36.23)	(-36.07)
Observations	1080	1078	1071	1069

Notes: t statistics in parentheses. All models include constant term and regional dummies.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 11: Heckman Endogenous Treatment Model Selection Equation

	(1)	(2)	(3)	(4)
English-medium				
Male	0.0407	0.0407	0.0407	0.0407
	(0.48)	(0.48)	(0.48)	(0.48)
Age	0.332*	0.332*	0.332*	0.332*
	(1.97)	(1.97)	(1.97)	(1.97)
Age^2	-0.00544	-0.00544	-0.00544	-0.00545
	(-1.94)	(-1.94)	(-1.94)	(-1.94)
High school				
Anadolu Voc.	-0.428	-0.428	-0.428	-0.428
	(-1.12)	(-1.11)	(-1.11)	(-1.09)
Vocational	-0.668*	-0.668*	-0.668*	-0.668*
	(-2.29)	(-2.28)	(-2.30)	(-2.29)
Science high	0.0317	0.0317	0.0317	0.0317
	(0.09)	(0.09)	(0.09)	(0.09)
General high	-0.413***	-0.413***	-0.413***	-0.413***
	(-3.73)	(-3.74)	(-3.74)	(-3.74)
Foreign lan/Sup.	-0.306*	-0.306*	-0.306*	-0.306*
	(-2.15)	(-2.15)	(-2.14)	(-2.14)
Previous English	0.486***	0.486***	0.486***	0.486***
	(4.67)	(4.66)	(4.70)	(4.70)
Mother worked	-0.282**	-0.282**	-0.282**	-0.282**
	(-2.75)	(-2.72)	(-2.75)	(-2.71)
Father Education (categoric)	–	✓	–	✓
Mother Education (categoric)	–	✓	–	✓
Observations	1080	1078	1071	1069

Notes: t statistics in parentheses. All models include constant term and regional (at high school) dummies.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$