Stockholm School of Economics
Department of Economics
659 Degree project in economics
Spring 2014

Is there a need for more male teachers in primary school?

A cross-country study on gender role model effects on educational outcome using PISA scores from 2000-2012

Caroline Nacksten ♀ Karolina Sjökvist ♀♀

Abstract

The Programme for International Student Assessments (PISA) has enabled cross-country studies on educational outcome, which is proven to be important for economic growth. The latest release of the 2012 PISA results has induced worldwide debates on how to best enhance the national outcomes. One of these debates is whether there is a lack of male role models in primary school given the general trend of an increased fraction of female teachers, and a concern for the result of male students. Is there a need for more male teachers in primary school? Previous studies on gender role models are mainly conducted as within-country studies and have found mixed evidence. This study attempts to contribute to these studies by investigating whether aggregate data supports a student–teacher gender interaction effect, and if this effect indicates that there is a gender role model effect that can explain cross-country differences in educational outcome. An education production function is estimated using a panel of 33 OECD countries and test scores from PISA between the years 2000-2012. We find no evidence of a need for more male teachers, but rather a positive effect of female teachers for both girls and boys, contradicting the gender role model theory. Nevertheless, the study finds strong indications for that student–teacher gender interaction is important in explaining differences in educational outcome.

♀ *22582@student.hhs.se* ♀♀ *22578@student.hhs.se*

Keywords: Student–teacher interaction, role models, male teachers, education production function, PISA
JEL classifications: I20, I21, I28, Z13

Supervisor: Yoichi Sugita
Date submitted: 14 May 2014
Date examined: 11 June 2014
Discussants: Alexis Lindhé and Nga Nguyen
Examiner: Örjan Sjöberg

Acknowledgements

First and foremost, we would like to thank our supervisor Yoichi Sugita for his commitment, support and helpful input. Furthermore, we wish to show our greatest gratitude to the teachers at the Department of Economics for providing valuable insights and tools for our academic development.

Table of contents

1 Introduction

The latest Programme of International Student Assessment (PISA) results have induced worldwide debates regarding the differences between countries in their educational outcome. For countries that were proven to perform below OECD average, such as Sweden, these results have induced great concern about their respective educational system. By looking at top performers, such as Finland, policy makers further try to find a best way to improve the educational outcome. The underlying reason for the attention drawn to standardized international student achievement tests, such as PISA, becomes evident in the educational production and economic growth literature, where these scores are commonly used as proxies for human capital formation, important for economic growth (Romer, 1990; Hanushek and Kimko, 2000; Barro and Lee, 2001). The results give countries an indication of the competitiveness of their educational system on the global arena. PISA, constituting the latest contribution of these test programmes, is argued to be better for international comparison than other similar tests, since PISA is designed to capture cognitive skills rather than curriculum-based skills (Hanushek, Woessmann, Schivardi and Pistaferri, 2011). Given that standardized international achievement test scores is a good proxy for human capital in growth studies, it becomes relevant to pin down the determinants of these scores (Hanushek and Kimko, 2000).

One of the debates that have arisen in light of the PISA results regards the increasing number of female teachers in primary school, commonly referred to as the feminisation of education. The high fraction of female teachers has been criticised for resulting in a lack of male role models. Boys are said to fall behind their female cohorts, and this lack of male role models is sometimes argued to be one explanation (Driessen, 2007). In line with role model theory, suggesting that a student who is assigned a same-gender teacher performs better in school due to a gender role model effect, the provision of more male role models will enhance the educational outcome for boys primarily (Dee, 2007). As of girls, given that they have the opportunity to be assigned a female teacher, adding the possibility of being assigned a male teacher can enhance girls´ performance as well, partly due to possible peer effects resulting from boys having a better attitude towards school (Hanushek and Woessmann, 2011).

In the literature on student–teacher interaction there is mixed evidence of the effect of a gender role model on student performance. Some claim that students assigned a teacher of the same gender will perform better, while others claim that the effect of gender of the teacher is dependent on the specific subject. The majority of studies of the student–teacher gender interaction are within-country studies and have used cross-sectional data.

The limited numbers of studies on student–teacher gender interaction together with the spurring debate of the gender composition of the teacher force, induce an interest in conducting a cross-country study to see whether a general gender role model effect on educational outcome prevails. And if so, whether the effect is of practical importance, relative to other commonly studied resource measures. The underlying purpose of our study is to provide guidance for policy makers on whether student–teacher gender interaction is of importance for educational outcome. More specifically, we seek to understand if there is reason to consider the composition of the teacher force, gender wise, resulting in our research question: Is there a need for more male teachers in primary school? We contribute to the studies on student–teacher interaction by examining the gender role model effect in a cross-country study using all five available PISA results. We do this by using aggregate data from primary school levels for OECD countries that have participated in at least two of the five PISA surveys. To study the effect of gender role models on educational outcome we estimate an education production function based on previous studies including a measure for the gender imbalance between students and teachers on an aggregate level. This study does not attempt to explain the gender gap in performance of boys and girls. Rather, it seeks to establish whether more male role models can enhance the aggregate score. This is based on the notion that male role models directly can enhance the performance of boys, which in turn improves the achievement of girls as a result of positive peer effects. Furthermore, this study does not attempt to distinguish among all determinants of educational outcome but rather to identify if there is support for a student–teacher gender interaction that can explain differences in educational outcome between countries.

2 Current state of knowledge

2.1 Education for economic growth

Much of the observed interest in educational outcome relates to the perceived importance of education to nurture future capabilities of students. A well-educated labour force is crucial for the economy. Endogenous growth theory suggests that human capital is important for economic growth (e.g. Lucas, 1988; Romer, 1990). Thus, many researchers try to find ways to assess its impact by finding suitable proxies. The quality of education in an economy directly affect a country's growth, not only because it is beneficial for the individuals participating but also because it provides a "rich environment for innovation, scientific discovery" which means that education can accelerate the growth rate (Hanushek, 2002, p. 2054). Current research has moved away its focus from quantitative measures of human capital towards a more qualitative viewpoint. Enrolment rates and quantity of schooling are broadly used proxies (e.g. Romer, 1990b; Barro, 1991; Mankiw et al., 1992) but have been criticized for not being adequate measures for educational quality. As an alternative to the quantitative measures, one can use direct measures of cognitive skills of individuals in order to permit quality to arise from factors outside of formal schooling (Hanushek and Kimko 2000). The most extensive literature on determinants of educational outcome use various standardized test scores, which in empirical studies have been found to be strongly related to economic growth (Hanushek and Kimko, 2000; Barro and Lee, 2001; Hanushek and Woessmann, 2009), hence motivated to use as a proxy for cognitive skills and thus human capital.

Existing studies utilizing cross-country variation in student achievement are either almost exclusively based on a cross-section of individual test performances in the same test of a single year, (e.g. Woessmann, 2003b; Jurges and Schneider, 2004) or the average in performance across many years (e.g. Hanushek and Kimko, 2000; Hanushek and Woessmann, 2007). The International Association for the Evaluation of Educational Achievement (IEA) programme conducted the first international test of academic achievement in 1964 and have then, together with IAEP programme, been the most commonly test scores used in cross-country studies (e.g. Hanushek and Kimko, 2000; Barro and Lee, 2001). These programmes, however, have been criticised for being curriculum based rather than capturing the abilities necessary for future challenges facing the labour force (Hanushek, Woessmann, Schivardi and Pistaferri, 2011). The OECD Programme for International Student Assessment (PISA) was first conducted in 2000 and have been increasingly used in favour of the previous tests since the survey is not curriculum-based but instead has a more real-world approach and assesses the skills that are considered to be essential for full participation in the society (OECD, 2014). Studies have used the extensive micro level data on a single year score or been weighted in the average of other test scores from other programmes in cross-country studies on the macro level. [1]

Math and science are the most commonly used subjects to look at when using international test scores as a measure of educational quality. Such concentration is consistent with the theoretical emphasis on the importance of research and development activities as the source of growth (e.g., Romer, 1990a) and has been empirically proved (e.g. Hanushek and Woessmann, 2007). Able students with a good understanding of mathematics and science is claimed to form a pool of future engineers and scientists. Reading has been excluded due to concerns about valid testing across languages and doubts about putting these scores into a common one-dimensional scale with science and mathematics tests (Hanushek and Kimko, 2000). Barro and Lee (2001) further found that the effect of inputs on reading scores are different from the effects of the same inputs on math and science scores at the same time as, for instance, family income were more important for reading scores.

Given our purpose to provide guidance on student–teacher gender interaction effects on educational outcome, international test scores seem to be a valid proxy for educational outcome in a cross-country comparison. To further achieve our aim, we

[1] For a review of studies using international achievement test scores, see Hanushek and Woessmann (2011).

look into the student–teacher interaction literature to find theories that can support why more male role models could be favourable.

2.2 Role model theory

Many scholars have found evidence that instructors act as role models, because they serve as examples to students or can better empathize with their particular needs (Bettinger and Long, 2005). Haveman and Wolfe (1995, p. 1834) define role models as:

"…adults or peers whom children or adolescents relate and who set norms of behavior and achievement to which they aspire…"

Studies on the effect of teacher gender on educational outcomes often take a standpoint in the main hypothesis that male and female teachers differ in how they engage girls and boys in the classroom (Dee, 2007). Students tend to respond to gender of their teacher, and not how the teacher actually behaves. Evans (1992, p. 211) defines a gender role model effect as:

"…an increase in the achievement of a female student if she has a female teacher, ceteris paribus."

Dee (2007, p. 532) further explains that a role model effect implies that:

"…a student will have improved intellectual engagement, conduct and academic performance when assigned to a same-gender teacher…"

Some researchers (e.g. Driessen, 2007) mention the feminisation of school, meaning that the percentage of female teachers constantly increases and that this means that there are fewer male role models for boys. Male role models might affect young boys' perception of masculinity and thereby change negative attitudes towards education and school.

The critique of the role-model effect is that the notion seldom is subject for critical evaluation, even though it could be understood from several different perspectives. One main criticism is that it is not certain whether girls and boys want to identify with teachers. Hence, making the issue a gender related question might induce problems since not all male teachers are suitable role models for boys, and not all female teachers are suitable role models for girls. Furthermore, the discussion of the increased amount of female teachers may not be completely accurate, since those managing, and thereby influencing, the schools most often are men (Driessen, 2007).

Alternative theories to the role model effect can be found in the social psychology literature where the existence of a stereotype threat and Pygmalion effect has been introduced. In educational sciences the stereotype threat has the implication that students who belong to a certain group associated with a negative stereotype will become more anxious in fear of confirming this negative stereotype. Hence, they will perform worse. This stereotype threat also affects the beliefs of students on what qualities girls and boys possess. This in turn, may affect the student outcomes (Antecol, 2012). The Pygmalion effect implies that the teachers' different expectations on boys and girls become self-fulfilling prophecies. Contrasting the psychological studies of gender differences, biological sciences provide theories of genetic differences between females and males. Tests have shown no difference in intelligence, but differences in scores of different cognitive tasks. Men outperform women in visual-spatial tasks, while women excel at certain verbal tasks (Dee, 2007).

Most studies on student–teacher gender interaction are mainly conducted using longitudinal student level data, or by experiments, in a single country or region. The studies we have reviewed have used micro level data, meaning that the individual score, attributes and the gender of the individual´s teacher have been assessed. Dee (2007) tests the educational outcome of girls and boys, separately, with a teacher in a certain subject as a function of observed student traits and a dummy for teacher gender. Antecol (2012) conducted a randomized experiment on the effect of teacher gender on student math achievement in primary school. His findings show that the female students who had female teachers performed worse than those with male teachers. When controlling for teacher background, this negative effect disappears if the teacher had a math-related major in college. In short, these findings imply that female students perform worse if their teacher is a female without a strong mathematical background. Dee (2007) further finds that the effect of having a female teacher falls if controlling for classroom and teacher variables. Boys who are assigned with a female teacher might suffer from lower scores. The results for girls are mixed and dependent on the subject. Evans (1992) does not find evidence for a gender role-model effect. Some claim that there is no strong effect on students learning outcomes when studying the match of teachers' race, gender and ethnicity and their students' equivalents. Instead, the matching of these characteristics appears to have a significant effect on the teachers' subjective evaluations of their students (Ehrenberg et al., 1995; Driessen, 2007). Driessen (2007) further finds no significant interaction effect on the number of male teachers on the competencies of primary school pupils. Bettinger and Long (2005) criticise the mixed results as they derive from a few number of studies, which in turn are case studies or limited studies. Despite the fact that many studies on teacher gender does not find a strong effect on student outcomes, Driessen (2007) claims that other arguments for attaining an equal gender balance for both teachers and managers in primary school could be found.

Previous research shows a need for looking at the student–teacher gender interaction effect in a broader context to see if there is an effect that can explain differences in educational outcome across countries. As previous research indicates, there is a need to control for classroom and teacher variables when trying to estimate the effect of student–teacher gender interaction. To find such a method, we turn to the literature on educational production.

2.3 Educational production function

The research on the determinants of educational outcome is rooted in production theory. The development of input-output analysis is said to originate from the Coleman Report, which has been influential among policy makers (Hanushek, 1986). The initial bivariate analysis, looking at the effect on a single resource on educational outcome, was criticised by Coleman who stated a need for multivariate analysis (Coleman et al., 1966). In an economic setting, the relationship measured by such input-output models is referred to as "educational production functions" and describes the maximum output that is feasible with different sets of inputs (Hanushek, 1979). Such an approach involve estimating some kind of function where the objective is to sort out the causal impacts of schooling factors from other influences on achievement, such as family background, peers etc. (Hanushek and Woessmann, 2011). However, in some older studies

An alternative to production theory is screening theory, which proposes that the social value of schooling may be smaller than the private value, if schools are only identifying the more able students rather than improving all students' skills. In a screening model, the output is information about the relative abilities of students, and this would suggest that more attention should be directed toward the distribution of observed educational outcomes (instead of simply the mean outcomes) and their relationship to the distribution of underlying abilities (Hanushek, 1979).

Hanushek (1979) presents a generally acceptable model, where student achievement (A) depends on social background (B), peer influences, (P), school inputs (S) and innate ability (I):

$$A = f(B, P, S, I)$$

Previous cross-country studies have shown that different levels of data can help explaining the difference in test scores across countries (e.g. Woessmann, 2003; Jurges and Schneider, 2004). However, the lack of data together with the fact that the multi-level of analysis is subject to endogeneity problems, force some researchers to focus on a single level (e.g. Hanushek and Kimko, 2000).

The early influential cross-country studies using aggregate data to explain differences in educational outcome has included family factors and school resources into their models (e.g. Hanushek and Kimko, 2000; Barro and Lee, 2001). The simplified model used in their work is specified as follows:

$$T = f(F, S)$$

where (*T)* is a function of family background (*F*) and school resources (*S*).

These factors are further justified by other empirical studies (e.g. Hanushek, Link and Woessmann, 2011). In these studies, the school resources are used to reflect the educational system present in the country. In some other studies there are country level factors other than school resources included. Hanushek and Woessmann (2011) include institutional structure into the model. Whereas the student outcome still is related to family and school factors, the productivity of these inputs now are affected by the institutional structure. However, it is evident in the literature that socio-economic factors as well as resources in both quantitative and qualitative terms are most frequently emphasized in explaining variation across countries.

The most frequently studied factors used in the education production functions are socio-economic (or family) factors and school resources. Below follows theories behind and empirical outcomes of these factors.

2.3.1 School resources

The effect of resources such as capital and labour on educational quality is the most extensive generally available evidence (Hanushek and Woessmann, 2011). The great deal of studies made on this factor is partly due to the availability of data, partly due to the possibility for policy makers to respond to the results of these studies (Hanushek, 1986). Educational spending, teacher salaries, teaching hours and pupil-teacher ratios are commonly used measures of such inputs. However, the evidence is mixed. In some studies, direct spending has been insignificant after controlling for GDP per capita (Hanushek and Woessmann, 2011) while being shown to be statistically significant in other country level empirical analysis (Barro and Lee, 2001). Teacher salary has further been shown to be an important driver for student achievement based on the theory that higher relative pay will provide teachers with more incentive for them to make more effort to improve the educational outcomes of the children they teach (Dolton and Marcenaro-Gutierrez, 2010). Teacher's salaries along with their education levels have been shown to be of great importance for student achievement (Barro and Lee, 2001; Jurges and Schneider, 2004; Dolton and Marcenaro-Gutierrez, 2010). Teaching hours is another dimension of schooling inputs that have been included in previous cross-country studies since more hours indicate how important education is perceived in a society. The effect has been proven positive and statistically significant (e.g., Barro and Lee, 2001; Alfonso and St.Aubyn, 2006). The pupil-teacher ratio is expected to have a negative effect on student achievement since students can learn more rapidly by having more frequent interactions with the teachers (Barro and Lee, 2001). Using the pupil-teacher ratio has had a positive effect on student outcome in some studies, i.e. when class size is smaller, (e.g. Barro and Lee, 2001) while being less significant in others (e.g. Hanushek and Kimko, 2000). Pupil-teacher ratio was not statistically significant when included together with teaching hours and country fixed-effect (Dolton and Marcenaro-Gutierrez, 2010).

Some macro level studies have not used any institutional factors in their models as resources are assumed to reflect the educational system (Hanushek and Kimko, 2000; Barro and Lee, 2000). Applying the theory of institutional economics to the schooling sector, it is the institutions of the education system that allocate the rights of decision-making in the system and determine the incentives faced by the actors (Bishop and Woessmann, 2004). Some control for institutions by including a dummy for central exit examinations, with mixed results (e.g. Bishop, 1997; Jurges and Schneider, 2004). However, the measure has been criticized for not capturing the cognitive skills but rather the test-taking ability (Piopiunik, Schwerdt and Woessmann, 2012). Other studies have used the degree of school decision-making autonomy in the country (Woessmann, 2003). Aggregated measures of the institutional feature can better capture the effect of private schools in a country. By comparing the average performance of systems with larger and smaller shares of private schools, the cross-country approach captures any systemic effect of competition from private schools (Hanushek and Woessmann, 2011).

2.3.2 Socio-economic factors

According to some studies, socio-economic factors are more important determinants of students' achievement than school resources (e.g. Hanushek and Kimko, 2000). Cross-country studies conducting analysis on the country level have found strong relationship between family background inputs and student achievement, measured by international achievement test scores, by using income and education of parents (e.g. Hanushek and Kimko, 2000; Barro and Lee, 2001; Jurges and Schneider, 2004; Alfonso and St. Aubyn, 2006).

Socio-economic background, measured by GDP, has been used as an indicator of overall skills and institutions. Higher-income countries tend to have better societal and economic institutions that promote productivity, societal vision, and smooth social interactions (Hanushek, Link and Woessmann, 2011). GDP per capita as a proxy for income has been commonly used in cross-country studies being statistically significant (Jurges and Schneider, 2004; Alfonso and St. Aubyn, 2006). In the regression of Barro and Lee (2001) GDP per capita was significantly negatively correlated with math and science scores when included in the cross-country macro level regression. Higher income also implies that the children have better nutrition and in turn an increased ability to learn. The reason for using measures for the education of parents is that highly educated parents tend to place a greater value on their children's education, and thus provide more materials and school-related activities. (Barro and Lee, 2001) Higher income would indicate better conditions for students to develop their cognitive skills while higher education reflect the perceived importance of education in the country. In some cross-country studies, population growth has also been included and has been shown negatively related to student achievement. An increased population implies that the same amounts of resources are to be allocated over a larger population (Hanushek and Kimko, 2000).

3 Research focus

In the student–teacher gender interaction literature, the studies have mainly been conducted using cross-sectional student level data within countries, and have focused on the gender gap. The evidence of the effect of being assigned a same-gender teacher is mixed but has been criticised for relying on few studies. The debate of the feminisation of primary school together with the fact that there are studies providing evidence of that there is a same-gender effect on educational outcome, motivate and indicate a need for further studies and clarification in the field. More specifically, there is a need for broadening the perspective and see if there is support for a general same-gender interaction effect in an internationally comparable context.

3.1 Limitations of scope

As opposed to other studies within the field of economics of education, we aim to find the student–teacher gender interaction effect rather than the effect of teacher gender in general on educational outcome. This way, we can capture whether there is an interaction effect and not only if female and male teachers seem to be better suited to teach different subjects.

Given the purpose of our study and limited availability of disaggregated data across countries, we limit the analysis to the country level in order to still get an indication whether there is support for a general student–teacher gender effect in primary school that explains differences in educational outcome across countries. In other words, this study does not try to assess the determinants of educational outcomes on a student or a gender base level, but rather on the aggregate country level. There is reason to believe that if we increase the fraction of male teachers, the outcome of boys will be improved due to the gender role model effect, and that this improvement will enhance the performance of girls as well due to peer effects and new influences from male teachers. Thus, a policy increasing the fraction of male teachers will affect both boys and girls and it is hence of interest to look at the aggregate outcome rather than outcomes by gender.

3.2 Research question

Given the limited studies conducted on student–teacher gender interaction in a broader, internationally comparable, context as well as with the current debate in mind, we identify a need for further assessing if there is a general same-gender interaction effect that can help explaining differences in educational outcome across countries. Given a gender role model effect, further presented in the following section, and given that female teachers generally dominate the teacher force, the research question is as follows:

4 Research design, methods and data

To answer our research question, we will use a quantitative research design and estimate an educational production function using aggregate data. To proceed with our study we formulate hypotheses, which will enable us to analyse and draw conclusions from our results.

4.1 Hypotheses

As a first step to provide an analytic framework for answering our research question, we need to investigate whether our aggregate data can support that a student–teacher gender interaction effect prevails. In other words, we aim to see whether a gender interaction between students and teachers can explain differences in educational outcome across countries. This hypothesis is based on previous research and theories claiming that the gender of teachers and students matter for educational outcome (Bettinger and Long, 2005; Dee, 2007; Antecol, 2012).

Hypothesis 1

A student–teacher gender interaction effect prevails.

The null hypothesis is that no effect can be identified, i.e. that there is not a statistically significant effect when estimating our education production function, which is specified in the following section, 4.2 Econometric specifications. If we can reject the null hypothesis, we can confirm that it is motivated to look at the composition of the teacher force, gender wise, since it can help explaining differences in educational outcome across countries. The practical significance is further of great interest and will be analysed as how large this interaction effect is relative to other commonly investigated school resource components.

Given that we can identify a statistically significant student–teacher gender effect, we further wish to investigate whether this student–teacher interaction effect has a sign that can indicate a gender role model effect. Hence, we can establish whether there is a need for more male role models in primary school. Consequently, our main hypothesis is based on previous research and theories that claim that a match between teacher and student gender plays a role in explaining educational outcome. More specifically, we rely on the notion, expressed by Dee (2007, p. 532), that a gender role model effect means that

"…a student will have improved intellectual engagement, conduct and academic performance when assigned to a same-gender teacher…"

This theory is the most dominant in the student–teacher interaction literature. As the lack of male role models is of interest in the current debate, it is further motivated to test whether this theory is supported in our cross-country study. On the country level of analysis, this match would be obtained if all girls and boys have equal opportunities to be assigned same-gender teachers. In other words, an imbalance in the provision of female and male teachers in relation to the composition of female and male students would have a negative effect on educational outcome.

Hypothesis 2:
A student–teacher gender imbalance has a negative effect on educational outcome.

4.2 Econometric specification

To test our hypothesis we estimate an educational production function model since it is the far most common approach when examining the determinants of educational outcome measured by international achievement test scores. The model used is justified in previous research (e.g. Hanushek and Kimko 2000; Barro and Lee, 2001) since it was designed for cross-country studies on the macro level. Theoretically, the educational production is expressed as follows:

$$T = F(S, R) + \varepsilon$$

Where *(T)* denotes educational outcome, *(S)* socio-economic factors, and *(R)* school resources.

Direct observations of educational outcome are available for more than 70 countries that have, at least once, participated in PISA during the five times that the survey has been conducted. However, the number of countries with available aggregate panel data is limited to 37 countries where three of these are still considered non-OECD countries (Russian Federation, China and Indonesia). One could extend the data set by using scores from other programmes, but since we believe that PISA is a better measure for international comparison, and since it allows for panel data, we choose to limit the data collection to PISA scores only.

We develop a panel of international test results from the Programme for International Student Assessment (PISA), covering 33[2] countries and five waves that span a time period of twelve years. It is an unbalanced panel since some countries have only participated three or four of the five test years[3] (Wooldridge, 2013). The panel character of the analysis is on the country level to see if the gender role model effect is present in a broader context, relating the effect to other commonly studied resources that are said to be important for educational outcome. By using aggregate data we can put the interaction effect in a context of standard and readily available resource measures.

[2] China, Indonesia, the Russian Federation and Turkey are dropped to avoid outliers.
[3] See Appendix: Table A.2 for an overview of the participating countries by each test year.

The downside of looking at the country level is that macro education production functions cannot control for individual influences on a student's performance (Woessmann, 2003). PISA provides an extensive database of student characteristics, but there is lack of data on teacher characteristics. However, by not using student-level data we do not have to pay attention to the complex data structure produced by the survey design or the multi-level nature of the explanatory variables. Also, since the international achievement tests do not follow the same individuals over years it is not possible to use panel data on this level. Furthermore, we assume that the innate ability of students and teachers are on average the same across countries, or at least exogenous to the other determinants accounted for in our model.[4] This enables us to focus on the population characteristics and compare the proposed gender role model effect to the effects of other commonly studied aggregate resource measures.

By using cross-country data we can capture variation that is not possible to capture in within-country data. Systematic institutional variation between countries as found with differences in educational spending and teacher salaries and the extent of a private school sector simply does not exist within most countries (Hanushek and Woessmann, 2011). The fractions of female teachers and female students are likely to be more constant within countries than across countries, thus we gain insight in the importance of a student–teacher gender interaction effect on education in a cross-country context. For a policy-making purpose, the key institutional factors, that are reflected in how many women and men are entering the teacher force and school, are possibly better captured in cross-country studies than in within-country studies (Hanushek and Woessmann, 2011) Thus, we can map if there is a general pattern of a same-gender interaction effect rather than merely a country dependent phenomena.

A drawback of using cross-country variation is that there may be unobserved factors that are hard to control for. In our cross-country study there is reason to believe that there are unobserved country-specific social and cultural factors that has an effect on our input variables as well as on the educational outcome. In our study these could reflect differences in how important education is perceived to be in the country and thus how much effort both teachers and students put into the education. In addition to the cultural differences, there is reason to believe that the level of difficulty on the tests differ between the test year due to the unreasonable ability to provide perfectly comparable tests in two years in terms of the level of difficulty.

Based on the need for controlling for such cultural factors and time-dependent factors the educational production function, relating test scores, $T_{i,t}$, to inputs in our panel of countries, can be specified as follows:

$$T_{i,t} = \beta_0 + \beta_1 ROLE_{i,t} + \beta_2 X_{i,t} + \alpha_i + \delta_t + u_{i,t}$$

[4] This reasoning is in line with previous studies (e.g Barro and Lee, 2001).

where i refers to the country of observation, t refers to the time period, β_i is the coefficient, i.e. the effect of the variable, α_i refers to country-specific unobserved factors, δ_t refers to the time specific unobserved factors, and $u_{i,t}$ is the idiosyncratic (or time-varying) error. $ROLE_{i,t}$ represents the main explanatory variable, which is further specified in section 4.4.3 Main explanatory variable. $X_{i,t}$ represents the control variables, consisting of different sets of socio-economic factors and school resources. These are further presented in section 4.4 Quantitative data. Our primary interest in estimating the model is the statistical significance, the sign and the magnitude of β_1. The partial effects of the control variables are of secondary interest, even though they can still be valuable for implications regarding policy-making decisions.

4.3 Econometric estimation

Our panel data set allows our parameters to rely on cross-country variation in the dependent variables across countries and over time.[5] Also, since we have a relatively short time period we can assume that the unobserved heterogeneity does not vary over time, which makes panel data preferable compared to cross-sectional data. Under the above conditions, fixed effects or random effects estimation will provide consistent estimates of the marginal relative importance of a gender role model effect on educational outcome (Wooldridge, 2013).

In both unobserved heterogeneity models we have a composite error term, $v_{i,t}$, that constitutes of an unobserved heterogeneity error, α_i, and an idiosyncratic error, $u_{i,t}$. Unobserved heterogeneity is the part of the error term that is fixed and independent of time. If the unobserved heterogeneity is not correlated with the independent variables, random effects estimation can be used for unbiased estimates. On the contrary, if we assume that the unobserved heterogeneity is correlated with the independent variables, fixed effects estimation is better since it allows for such correlation without causing biased estimates (Wooldridge, 2013). In our case, the unobserved heterogeneity, such as social and cultural factors, is likely to be correlated to the explanatory variables in our regression. The importance of education perceived in the country may induce different level of efforts and resources spent on education. Thus, in order to receive consistent estimates fixed effects estimation is motivated (Wooldridge, 2013). Another way to solve the problem with endogeneity would be to use the instrumental variable approach. However, just like other researchers (e.g. Barro and Lee, 2001) we have not found a good instrumental variable available at the aggregate level. Hence, the fixed effect approach is used for our study.

Given our OECD sample, the countries are similar in the level of development, which is likely to affect the culture that prevails. In other words, the heterogeneity is reduced. This indicates a smaller need for controlling for country-specific factors than

[5] See Appendix: TABLE A.2.

if studying both poor and rich countries.[6] Instead we use region fixed effects[7], using regions that are likely to have similar cultures. The geographical regions are the suggested ones for statistical analysis according to the United Nations Statistics Division (2013).[8]

In addition to the cultural differences, there is reason to believe that the level of difficulty on the tests differ between the test year due to the unreasonable ability to provide perfectly comparable tests in two years in terms of the level of difficulty. Therefore, we control for year fixed effects as well. When we have eliminated the fixed effect before the estimation we can estimate the effect by the fixed effects estimator, Ordinary Least Squares (OLS), which is commonly used and suggested for analysing production functions with a single outcome (Hanushek, 1979). A problem with OLS is that it is sensitive to outliers due to the fact that it minimizes the sum of squared residuals. For this reason, we drop the observations for Turkey, Indonesia, Russia and China to receive unbiased estimates (Wooldridge, 2013). Country-level regressions of student achievement are particularly sensitive. Hence, there is reason for interpreting cross-country evidence with some caution (Jurges and Schneider, 2004).

The key assumption that we have to make is that the idiosyncratic errors are uncorrelated with the explanatory variable in each time period. Furthermore, since the variances of errors have to be homoscedastic for a reliable estimate, standard robust errors are employed in all our regressions. Also, the idiosyncratic error terms are assumed to be serially uncorrelated, i.e. the idiosyncratic errors in two different time periods are uncorrelated given the independent variables (Wooldridge, 2013). The idiosyncratic errors are further assumed to be independent of our explanatory variables and normally distributed, enabling us to use t-statistics for significance evaluation.

Because of the low evidence of lagging effect on the estimated effects of inputs in a macro educational production function along with support from previous studies (Barro and Lee, 2001; Dolton and Marcenaro-Gutierrez, 2010), we match our input measures to the test score in the same year, or the nearest available input measure if data is not available for the test year in question.[9] Resource measures, such as pupil-teacher ratio and the expenditures, are most likely exogenously determined (at least in the short run). This is because the decisions are most commonly made based on

[6] Using country-fixed effect would make us lose much information, and the country fixed effects are likely to be omitted due to multicollinearity since they are so similar in terms of educational systems and culture.

[7] Region fixed effect is use in previous studies as well (e.g. Hanushek and Kimko, 2000). Some studies include a dummy for East Asian countries only (e.g. Barro and Lee, 2000), since these are assumed and proven to differ substantially in how important education is perceived. However, we claim that it is important to control for less extreme cultural differences as well.

[8] See Appendix: Table A.2 for the representation of regions. In our sample, some regions are under-represented in relation to the other regions. This could make the regional estimates misleading. South America and Central America, which can be assumed to have a similar culture, are clustered to decrease this unbalanced region classification.

[9] Variables in years prior to the test year have been preferred over variables in consequent years due to possible reverse causality. If the variables have not been available for two years before or after the test year, the data has been marked missing.

administrative conventions and political directives from government, at the level of the individual country (Dolton, 2010). Endogeneity problems are further logically less severe in cross-country studies than in within-country studies (Barro and Lee, 2001). There is no mechanism that would distribute resources across countries as a result of differences in educational outcome. Also, even if mobility across countries is easier today than before, it is not likely that people move across countries due to the student achievement in the country.

4.4 Quantitative data

The aggregate data is retrieved from the broadly accepted independent statistical databases supplied by UNESCO and OECD. For an overview of the descriptive statistics of the variables used, see TABLE 1. For data sources and variable specification, see TABLE A.1 in Appendix.

4.4.1 Dependent variable

Given that standardized international student achievement test scores have been proven to be important in human capital formation in growth studies (e.g. Hanushek and Kimko, 2000; Barro and Lee 2001), using such a score as our dependent variable serves as a good proxy for educational outcome. Due to the favourable attributes of PISA[10], both for its ability to capture skills important for economic growth and for its comparability across countries, we choose to use PISA scores in math and science[11] as our dependent variable. Reading is excluded due to difficulties of comparing these scores across countries. There is also reason to believe that reading skills have other main determinants than math and science skills. This pattern has been proven by previous studies (Barro and Lee, 2001). Furthermore, there is more evidence on math and science being important for growth (e.g. Hanushek and Kimko, 2000), which further motivates our choice of dependent variable.

As the PISA score is measuring cognitive skills[12] (i.e. problem solving abilities) rather than curriculum-based skills, we limit ourselves to control for primary school data. One can intuitively argue that this level provides the foundation for future academic and life-long learning. Even though PISA is measured for 15-year olds who attend lower secondary school, we argue that by using primary school data for the school related measures we can assess the important drivers for educational outcome in the country. Moreover, previous studies have failed to find significance including lower secondary levels (e.g. Hanushek and Kimko, 2000; Barro and Lee, 2001).

---

[10] PISA assesses how well 15-year old students have acquired some of the knowledge and skills that are essential for full participation in society (OECD, 2014).

[11] We take the country mean PISA scores in math and science and then average these. For the United Kingdom scores for Scotland and England have been averaged. For Belgium, the scores for Belgium Fr. and Belgium Fl. have been averaged.

[12] Due to availability of data, we are forced to narrow our analysis to cognitive skills, measured by PISA. This induces concerns about non-cognitive skills that cannot be accounted for.

4.4.2 Main explanatory variable

In order to operationalize our hypotheses we construct our own measure that we believe can reflect the student–teacher gender imbalance in a country. As previously mentioned, the main hypothesis is based on the notion that teachers serve as role models, claimed by several researchers (e.g. Evans, 1992; Bettinger and Long, 2005), and that being assigned a same-gender teacher improves the intellectual engagement, conduct and academic performance of the students (Dee, 2007). This optimal classroom effect is, on the aggregate level, theoretically achieved when the teacher force represents the student force in a gender composition aspect.[13] A balanced gender distribution between teachers and students is beneficial for the aggregate educational outcome. To measure the level of imbalance, the main explanatory variable is expressed as follows:

$$ROLE_{i,t} = \left| Fraction\ of\ female\ teachers_{i,t} - Fraction\ of\ female\ students_{i,t} \right|$$

In words, the measure denoted ROLE[14] in our regressions, is the absolute difference between the fraction of female teachers and the fraction of female students to reflects the imbalance between the teacher and student gender composition on an aggregate level. An imbalance would imply that a particular student gender has a higher probability of interacting with same-gender teachers. In our sample, the country fraction of female teachers is always higher than the fraction of female students. This indicates that the larger the difference, the higher the lack of role models for male students. An imbalance in the gender composition between teachers and students in primary school imply that there are more female teachers than optimal, relative to the fraction of female students. If we find a significant estimated effect of ROLE, we get support for hypothesis 1. Moreover, if this effect is negative, we get an indication of that a gender role model effect prevails. In other words, we thus get support for hypothesis 2.

4.4.3 Control variables

In order to capture the proposed effect of ROLE, we need to control for other factors affecting educational outcome and that are correlated with ROLE. These factors are based on previous research and theory of educational determinants. Measures of absolute values, such as GDP per capita and salaries are presented in 2000 international dollars for comparability across countries and over time.[15]

[13] We assume that the fraction of female teachers and female students are representative for all schools in a given country. This is a strong assumption, but we believe that this simplification of reality still serves a good enough proxy for finding an indication on whether there is a same-gender effect. We further assume that the fraction of female teachers is representative for all subjects. Even though there is a possibility that male and female teachers end up teaching different subjects in general, looking at the primary education reasonably makes this problem less severe since children are more likely to have the same teacher in all subjects to a greater extent in primary school than in higher education.

[14] ROLE is expressed in absolute terms. In our dataset though, the difference is never negative as the fraction of female teachers is always larger than the fraction of female students.

[15] The GDP measures have been presented in current dollars and current PPPs (purchasing power parity index). For comparison purposes we have used the GDP deflator to convert the data into base year 2000. Teacher salaries were reported in current dollars and current PPPs, and have consequently been CPI (consumer price index) adjusted to base year 2000.

Socio-Economic Factors

GDP per capita: GDP per capita is included as a proxy for the social structure and standard of living in the country as high income should provide a more stable social interaction, beneficial in the classroom. Even though there is mixed evidence, we expect a positive effect of income on the test score according to theory.

Adult education: To control for the perceived importance of education in the country we include the percentage of the population who has attained a tertiary degree. Educational attainment is closely related to the skills and competencies of a country's population, and is thus an important factor in human capital formulation. At the same time this measure can serve as a proxy for the education of teachers. According to theory and previous studies, we expect a positive effect of adult education on test scores.

Population growth: Population growth is included to control for the trade-off between quality and quantity. As in previous studies we expect the population growth to be negatively related to the mean test score since higher growth means less qualitative attention and resources to education.

School Resources

Government expenditure per student: We include government expenditure per student (as a percentage of GDP per capita) to control for the allocation of resources to each student in the country. This measure is of interest for policy makers and has thus been commonly included in educational production functions. Even though there is mixed evidence on whether expenditure matters in the presence of other, and perhaps better, resource measures, we refer to production theory as many previous studies have, and expect expenditure to be positively related to the mean test score.

Teacher salaries: Teacher starting salaries have been included to control for teacher quality, as higher salaries tend to attract more able teachers. As previous studies have pointed out, this measure may not be ideal for capturing teacher quality and that teacher background controls should be favoured. In primary levels of education, the specific subject knowledge requirements of teachers is not of the same importance as it is in higher levels of education, and therefore the starting salaries could be argued to sufficiently capture the teacher quality. Thus, we expect a positive sign of teacher salary.

Pupil-teacher ratio: The pupil-teacher ratio is included as a proxy for classroom resources as smaller classes enable more interaction between teachers and students. Even though the evidence is mixed, it is included for the same reason as in other studies. Primarily based on theory, we expect a negative impact of the pupil-teacher ratio on the test scores.

Teaching hours: Teaching hours are included in the regression to proxy for how many hours students in a country interact with their teachers per year. The more hours, the more time and pedagogical resources to learn and we therefore expect a positive effect of the number of hours on test scores is expected.

Fraction of private schools: To control for institutional factors other than those captured by our other resource measures we include the fraction of private schools. We expect the sign to be positive, since more autonomy leads to higher efficiency and quality.

Fraction of female teachers: We add FEMT as an additional control variable for female teacher quality. According to previous studies, there is reason to believe that female teachers have teaching qualities that have a positive effect on educational outcome, and that may be captured by ROLE. If there is such an effect, by controlling for FEMT we see if there is still a gender interaction effect given a certain level of female teachers in the country. In other words, the characteristics for female teachers that are not related to ROLE will be captured by this fraction instead. Given a positive effect of FEMT we expect a more negative effect of ROLE. We are aware of the high correlation of FEMT and ROLE, as FEMT is a part of ROLE. However, due to the lack of data on teacher quality, we include FEMT to get an indication of the effect of ROLE after controlling for such factors.

4.5 Data issues

PISA is good in the sense that a country cannot easily withdraw from reporting the test scores while they have participated. Thus, the result of ill performing countries will still be reported (PISA 2012 results, OECD). However, the PISA scores may be subject to non-random sampling since the participating countries themselves have to apply for, finance and administer the test. This implies that the selection of countries is not fully random as the cost of financing and administration requirement might hinder poorer countries to participate. However, there are not only OECD countries participating. In the latest PISA survey, 80 % of the world economy was represented (OECD, 2014). Since we look at OECD countries, we mitigate this problem of non-random sampling. However, we cannot apply our results in a context outside OECD or advanced economies.

For the sample as a whole, we have missing data on some variables for some countries that make us lose some degrees of freedom, making our estimates less precise. The missing data will cause sample selection problems if it is correlated with the idiosyncratic errors (u_{it}) (Wooldridge, 2013). In our sample, the missing data is mainly random since there is lacking data on a certain year for a certain variable. However, for some countries data on certain variables have not been reported at all. Thus, one has to be aware of the distortion effects that this could result in.

A difficulty in any study of this sort is the existence of heterogeneity in the educational systems of the different countries that cannot be easily observed. When reporting data on this aggregate level, it is difficult to ensure that the data is based on the same classifications. For example, EDUC measures the fraction of the population that has a tertiary degree. The tertiary degree is classified as ISCED level 5 and 6.[16] These levels are categorised for international comparability. However, since countries have different educational systems the countries may experience difficulties in categorising their educational levels into the ISCED classifications. Some measurement problems may thus come as a result.

The fact that some of our resource measures are for both public and private, while others are for public only, further needs to be noticed. These differences may not distort our results. Rather, it induces a need for caution when interpreting the specific coefficients for policy-making implications.

In the gender role model measure, the largest variation is found in the fraction of female teachers while the fraction of female students does not vary considerably over countries or time. Thus, the fraction of female teachers is the main driver for ROLE. This is important to have in mind when interpreting the findings.

4.6 Regressions

In order to identify whether a general gender role model effect prevails, we run different regressions. First, we run three regressions to test the validity of our education production function with our panel data set. The composition of control variables is in line with the study of Barro and Lee (2001). To analyse the validity of the model, we compare the results to their findings and other previous studies conducted on an aggregate level. This regression, later referred to as the "base regression", will be run with and without region and year fixed effect (hereafter referred to as FE) as well as with an East Asian dummy in accordance with the study by Barro and Lee (2001).

$$PISA_{i,t} = \beta_0 + \beta_1 LOG_{GDP_{i,t}} + \beta_2 EDUC_{i,t} + \beta_3 EXP_{i,t} + \beta_4 LOG_{SAL_{i,t}} + \beta_5 PTR_{i,t} + \beta_6 HOURS_{i,t} + \alpha_i + \delta_t + u_{i,t}$$

Second, we add ROLE to the base regression. This will be referred to as the "main regression" and is specified as follows:

$$PISA_{i,t} = \beta_0 + \beta_1 LOG_{GDP_{i,t}} + \beta_2 EDUC_{i,t} + \beta_3 EXP_{i,t} + \beta_4 LOG_{SAL_{i,t}} + \beta_5 PTR_{i,t} + \beta_6 HOURS_{i,t} + \beta_7 ROLE_{i,t} + \alpha_i + \delta_t + u_{i,t}$$

[16] ISCED: *International Standard Classification of Education* is a classification of educational levels developed by UNESCO for facilitating cross-country comparisons of education statistics and indicators (UNESCO, 2014).

Third, with the main regression as our starting point, eight regressions with different compositions of control variables will be run in order to identify the proposed effect of ROLE on PISA. This is to make sure that over controlling or omitted variable bias do not distort our estimate of ROLE. For guidance on these effects we use the correlation of the variables in our data set.[17]

Regression (M1) and (M2) are the main regression with and without FE. We expect the effect of ROLE to be more significant but with lower magnitude since cultural factors are accounted for and are likely to affect the quality of both student and teacher efforts. The social interaction is likely to be affected by the culture that prevails. Regression (M3) to (M5) show the results of dropping one explanatory variable at a time from the main regression in order to see how much the model explains with different compositions to understand how ROLE is affected. The variables are dropped, either due to the risk of multicollinearity or due to the mixed evidence on their significance and in education production functions. In regression (M6) we will control for PRIVATE as more autonomy would lead to more classroom effectiveness and by not controlling for PRIVATE this effect could be captured in ROLE, and consequently could distort its coefficient due to omitted variable bias. Since the correlation is negative between ROLE and PRIVATE, and that we expect a positive effect of PRIVATE. By not controlling for PRIVATE we underestimate the effect of ROLE. Thus, we expect ROLE to increase in this regression. In regression (M7) we add POPG to capture the impact of an increased population and the lowered resources associated with it, both in classrooms and in terms of expenditures. As ROLE and POPG are negatively correlated and POPG assumedly negatively related to PISA we expect ROLE to decrease when controlling for POPG. The last regression (M8), FEMT is included. As FEMT and ROLE are near perfectly correlated we do not expect these estimates to be significant. Rather, we expect to get an indication on whether there is a gender role model effect after controlling for female teacher characteristics.

4.7 Robustness

To test the robustness of our results we first test the main regression by using average[18] of primary and lower secondary data since the test is taken by 15-year-olds indicating that the lower secondary may influence the educational outcome. Further, we run the main regression with the primary data again, now using European countries only since they constituted the main part of the full sample. This way we get an indication if the results hold in another context. In the third robustness test regression (R3) we use an East Asia dummy in line with previous studies as East Asia is said to differ the most in terms of perceived importance of education as well as in

[17] See Appendix: TABLE A.3

[18] The average is taken since we thus get the average resources endowed during the compulsory school. Including lower secondary data separately together with primary school data, may induce endogeneity problems since salary, for example, in lower secondary school is likely to be correlated with salary in primary school in a given country.

terms of the structure of the educational system. As a last robustness test we decompose the ROLE measure and see if the net effect of FEMT and FEMS differs from the effect of ROLE.

4.8 Descriptive statistics

Below we present descriptive statistics of our data set.

TABLE 1: DESCRIPTIVE STATISTICS

Variables	Observations	Min.	Mean	Max.	Std. Dev.
PISA	152	395.0500	499.3050	555.8406	30.4660
LOG_GDP	165	9.1637	10.1145	11.0874	0.4123
EDUC (%)	158	8.8400	27.4428	51.3200	9.7432
EXP (%)	147	10.5692	20.5177	31.1325	4.2204
LOG_SAL	138	8.5366	9.9802	10.8354	0.4563
PTR	132	9.2663	15.4558	34.0928	5.2377
HOURS	138	583.0000	797.2631	1139.4000	129.4985
ROLE	131	8.4092	33.0069	49.1399	9.1425
FEMT (%)	131	56.8283	81.5419	97.6296	9.1162
POPG (%)	165	-0.7892	0.6254	2.8628	0.6597
PRIVATE (%)	157	0.0000	10.6454	68.9125	15.0460

The variation in all our variables is considered sufficiently high for retrieving results, from which we can draw conclusions. There are differences across countries in PISA, implying that there is motive for pinning down the determinants for these differences. Furthermore, we identify variation in ROLE, which mainly is driven by differences in the fraction of female teachers. The fraction of ROLE is rather stable across years for a given country. However, small variations are still present. For example, in Greece ROLE has gone from 8.41 to 16.79 between 2000 and 2009. There is further little variation across countries and years in the fraction of female students but still differences that affect ROLE and thus of interest to take into account when studying the gender interaction effect.[19] See Appendix for more detailed data description.

5 Empirical Results

5.1 Main regressions

The regression explaining the most, i.e. the highest adjusted R^2, is the regression where POPG is included. The significant estimates in this regression are EDUC, LOG_SAL and ROLE. In all regressions ROLE is significantly positive, except when controlling for FEMT where it is negative and insignificant. The positive coefficient ranges from 0.5312 to 1.5528, whereas the negative coefficient is -0.4231. This means that an increase in ROLE decreases PISA by 0.4231 points. The ROLE estimate loses some of its significance when LOG_GDP and LOG_SAL are dropped.

When controlling for FE, the size of the estimates falls. When dropping LOG_GDP, ROLE becomes slightly less positive. When controlling for PRIVATE the model explains more, and ROLE decreases slightly, while staying statistically significant at the 1 % significance level. Adding POPG causes ROLE to decrease slightly and is now statically significant only at the 5 % level. Adding FEMT makes ROLE insignificant and negative. The standard deviation of ROLE in this regression became substantially larger, 6.6782, than compared to the other regressions where the standard deviation ranges from 0.2155 to 0.3697.

EDUC is always positive and statistically significant at the 1 % level, and does not change notably in the different regressions. LOG_GDP is positive and insignificant in the regressions without FE, while negative and significant in the others (with an exception for the regression when dropping LOG_SAL). EXP is always negative but only significant without FE. The standard deviation of EXP is large in relation to the coefficient. Thus, in some cases EXP can have a positive impact on PISA. LOG_SAL is always positively significant and ranges from 18.5061 to 22.9432. The significance increases when using FE. Dropping LOG_GDP makes it insignificant and substantially smaller. The coefficient for PTR is always positive but never statistically significant. This implies that the larger the class size, or the more students per teacher, the better the educational outcome. The standard deviation, however, is often larger than the coefficient, implying that the effect of PTR on PISA can be negative. HOURS is only significant without FE and the coefficient is negative in all regressions The standard deviation is low compared to the coefficient. The coefficient for FEMT is insignificantly positive, the coefficient for POPG is significantly negative at the 10 % level and PRIVATE is insignificantly negative.

[19] In our sample of OECD countries, boys and girls can be assumed to have the same opportunity to attend school. Thus, the little variation still found in FEMS is rather a reflection of natural reasons such as the difference in the number of boys and girls born. However, regardless of the reason it still becomes important to take this variation into account when studying the gender interaction effect.

TABLE 2: MAIN REGRESSIONS

	Dependent variable: PISA							
Independent variables	(M1)	(M2)	(M3)	(M4)	(M5)	(M6)	(M7)	(M8)
LOG_GDP	15.2269	-23.2617***		-7.1067	-23.9679***	7.1308	-9.9880	-24.2705**
	(12.0356)	*(8.6587)*		(7.2024)	(7.7390)	(11.9781)	(11.4583)	(11.2619)
EDUC	1.2929***	1.1296***	1.2093***	1.1674***	1.1546***	1.661***	1.2839***	1.1101***
	(0.4172)	*(0.3096)*	(0.3196)	(0.3248)	(0.2974)	(0.3397)	(0.3126)	(0.3684)
EXP	-1.4426*	-0.8411	-0.6351	-1.0219	-0.4611	-1.7522**	-0.7817	-0.8458
	(0.8326)	*(0.6750)*	(0.7250)	(0.6799)	(0.5721)	(0.8054)	(0.7131)	(0.6778)
LOG_SAL	18.8828*	22.5343***	8.3841		19.6936***	20.3956**	18.5061**	22.9432***
	(9.6541)	*(7.7317)*	(6.9917)		(7.0587)	(9.7005)	(7.5643)	(7.7655)
PTR	1.0049	0.3814	0.6945	0.3365	0.4074	0.4402	0.7747	0.3649
	(0.9351)	*(0.6513)*	(0.6352)	(0.6772)	(0.5950)	(0.9737)	(0.7965)	(0.6687)
HOURS	-0.1334***	-0.032	-0.0317	-0.0285		-0.1279***	-0.0273	-0.0305
	(0.0198)	*(0.0228)*	(0.2503)	(0.0235)		(0.0206)	(0.0212)	(0.0250)
ROLE	**1.5528*** **	**0.6525*** **	**0.6375** **	**0.5312** **	**0.5792*** **	**1.5471*** **	**0.5825** **	**-0.4231**
	(0.3685)	***(0.2376)***	**(0.2503)**	**(0.2363)**	**(0.2155)**	**(0.3697)**	**(0.2488)**	**(6.6783)**
FEMT								1.0788
								(6.6529)
POPG							-7.8864*	
							(4.5127)	
PRIVATE						-0.1934		
						(0.1209)		
Constant	192.6844**	509.3747***	403.1581***	572.9658***	515.0315***	263.7419***	405.7667***	462.7027*
	(95.1775)	*(93.0912)*	(82.3048)	(85.0692)	(85.0929)	(94.8720)	(111.6308)	(274.6918)
FE	-	Yes	Yes	Yes	Yes	-	Yes	Yes
No. of obs.	89	89	89	94	98	88	89	89
R^2	0.4425	0.8638	0.8585	0.8569	0.8548	0.494	0.8769	0.8674
Adjusted R^2	0.3943	0.8263	0.8221	0.8226	0.8217	0.4428	0.8407	0.8284

*Notes: Robust standard errors in parentheses. ***Significant at 1 % level (p<0.01), **Significant at 5 % level (p<0.05), *Significant at 10 % (p<0.1)*

5.2 Robustness regressions

Using average data from primary and lower secondary school in our main regression with and without fixed effect, the significance is completely lost for all variables except LOG_GDP and EDUC. LOG_GDP is then statistically significant at the 10 % level while EDUC stays significant at the 1 % level. Remarkably, ROLE loses all significance, but remains positive. In the regression using a subsample of only the European countries, ROLE becomes insignificant and decreases in size. PTR becomes significant for the first time but remains positive. LOG_SAL becomes slightly more positive, and stays equally significant. In the regression with the East Asia dummy, ROLE becomes substantially higher than in the main regression with FE. LOG_GDP and LOG_SAL become insignificant. PTR here has a negative coefficient and is significant at the 5 % level. When decomposing ROLE into FEMT and FEMS, the other control variables do not differ substantially from the main regression using FE. The coefficient for FEMT is significantly positive at the 1 % level, whereas FEMS is insignificant and positive. The coefficient for ROLE in the main regression is quite similar to the coefficient for FEMT in the robustness regression.

TABLE 3: ROBUSTNESS REGRESSIONS

Independent variables	Dependent variable: PISA					
	(M1)	(M2)	(R1)	(R2)	(R3)	(R4)
LOG_GDP	15.2269	-23.2617***	-20.3435*	-33.3609	13.6323	-24.2705**
	(12.0356)	*(8.6587)*	(11.3677)	(9.5677)	(10.3563)	(11.2619)
EDUC	1.2929***	1.1296***	0.9672***	0.456	1.1009***	1.1101***
	(0.4172)	*(0.3096)*	(0.3208)	(0.3481)	(0.4045)	(0.3684)
EXP	-1.4426*	-0.8411	-0.2353	-0.9326	-2.5243***	-0.8458
	(0.8326)	*(0.6750)*	(0.7973)	(0.6962)	(0.7275)	(0.6778)
LOG_SAL	18.8828*	22.5343***	12.2496	29.6536***	11.8988	22.9432***
	(9.6541)	*(7.7317)*	(8.5669)	(8.8713)	(8.2613)	(7.7655)
PTR	1.0049	0.3814	-0.1243	2.0593***	-1.4236**	0.3649
	(0.9351)	*(0.6513)*	(0.9577)	(0.7719)	(0.6867)	(0.6687)
HOURS	-0.1334***	-0.032	-0.0132	-0.0870***	-0.0972***	-0.0305
	(0.0198)	*(0.0228)*	(0.0331)	(0.02922)	(0.0216)	(0.02450)
ROLE	**1.5528*** **	**0.6525*** **	**0.5278**	**0.1787**	**1.5485*** **	
	(0.3685)	***(0.2376)***	**(0.2752)**	**(0.2920)**	**(0.2909)**	
FEMT						0.6557***
						(0.2377)
FEMS						0.4232
						(6.6783)
EAST ASIA					66.4881***	
					(10.3423)	
Constant	192.6844**	509.3747***	574.7213***	589.506***	308.77***	462.7027*
	(95.1775)	*(93.0912)*	(98.5284)	(83.6539)	(75.8148)	(274.6918)
FE	-	Yes	Yes	Yes	-	Yes
No. of obs.	89	89	90	67	89	89
R^2	0.4425	0.8638	0.8485	0.6221	0.5948	0.8674
Adjusted R^2	0.3943	0.8262	0.8101	0.4804	0.4906	0.8333

*Notes: Robust standard errors in parentheses. ***Significant at 1 % level (p<0.01), **Significant at 5 % level (p<0.05), *Significant at 10 % (p<0.1)*

6 Discussion

Firstly, we discuss our findings when testing the basic regression. Secondly, we discuss more in detail what seems to change the effect of ROLE in our regressions. Thirdly, we proceed by discussing the general results of a gender role model effect and implications of these results.

6.1 Regressions

Regression B1 to B3

When testing the model used in Barro and Lee (2001) with our data set, the result becomes similar but somewhat different in certain aspects.[20] In contrast to Barro and Lee (2001) we get a negative effect of LOG_GDP on test scores when using fixed effect. In previous research there is mixed evidence regarding the sign of expenditure, pupil-teacher ratio and instruction time. Comparing our results to these previous findings we see the same tendencies. Adult education is statistically significant and positive regardless of using fixed effect or not, in line with previous studies. One reason for the different results is that we have another sample of countries. Further we use a different data set of both the dependent variable and the independent variables.[21] To conclude, the results from using our estimated model with our data set indicate that the model is consistent with the results from previous empirical findings. Despite the, sometimes, contradicting signs of our variables, they are all considered important for studying the effect of student–teacher interaction on the aggregate level.

Regression M3 to M5

As mentioned before, regression M3 to M5 are mainly conducted due to the mixed evidence of some of the control variables. Our results show that the control variables are all important to include in our model since the adjusted R^2 decreases when excluding them.

When running FE, the size of ROLE decreases. This is in line with our expectation that constant cultural factors, such as the perceived importance of school, and year factors are important to control for as these seem to be captured by ROLE otherwise. When dropping LOG_GDP, ROLE becomes slightly less positive and the model explains less indicating that LOG_GDP is important and has an impact on ROLE. By not controlling for the social interaction captured by GDP per capita, ROLE will capture interaction effects not related to gender. Dropping LOG_SAL the estimate of ROLE becomes less positive and the model explains slightly less. Also, much of the

[20] See Appendix: TABLE A.4

[21] We use PISA instead of the test programmes conducted by IAE and IAEP, as Barro and Lee (2001) did. They use expenditures per student in absolute terms whereas we use expenditure per student as a fraction of GDP per capita for understanding the size of expenditures in relation to the income level of the country. We use teaching hours instead of length of school days to better capture the interaction time. For adult education we include the fraction of the population with a tertiary degree rather than the average school attainment since we have a sample of countries where the higher levels of education is what determines the differences in the perceived importance of education.

significance of the other variables is lost. As expected, the salaries are still important to control for since they capture teacher qualities that are unrelated to gender. Dropping HOURS decreases the estimate for ROLE and the model explains slightly less, indicating that we gain from including the number of instruction hours per year. The HOURS variable indicates the quantity of interaction between teachers and students. In our regressions, HOURS has a negative partial effect on educational outcome, contradicting the expectation. Since HOURS and ROLE are positively correlated we thus underestimate ROLE if omitting HOURS.

Regression M6 to M8

When adding PRIVATE, the estimate of ROLE does not change considerably in size or significance. This implies that the differences in institutions across countries reflected by PRIVATE do not have a large impact on ROLE. When controlling for POPG, the model explains the most. As expected, the effect of ROLE is less due to the supposedly negative social interaction effect associated with a higher population growth. If the partial effect of ROLE had been negative as expected, controlling for POPG would have resulted in a more negative effect of a gender imbalance in primary school. The negative effect of ROLE on educational outcome received when controlling for FEMT supports our hypothesis that there is a gender role model effect on educational outcome, i.e. an unequal gender balance between teachers and students affects educational outcome negatively. The positive effect of FEMT further implies that the positive effect of ROLE, previously discussed to reflect that females are better role models generally, may rather be due to that females have other qualities than male teachers that are beneficial for both boys and girls. This finding is in line with previous studies claiming that females may be better suited for teaching in general due to biological reasons (Dee, 2007). Even though FEMT in this context is not an optimal measure for female teacher characteristics as it is nearly perfectly correlated with ROLE, we get an indication that there are differences between female and male characteristics that are important to control for to get the real effect of a pure gender role model effect. However, we cannot determine with certainty whether the lack of significance depends on the ill-fitted measure of FEMT or that there is no pure role model effect after controlling for female teacher characteristics.

6.2 The effect of gender role models

The statistically significant effect of ROLE supports our first hypothesis that a student–teacher gender interaction effect prevails. However, the positive sign if this effect contradicts our second hypothesis, suggesting that an unequal gender balance between teachers and students has a negative effect on educational outcome. It contradicts the notion that girls and boys in primary school need a teacher of the same gender as a role model to improve their intellectual engagement and conduct in order to enhance their academic performance (Dee, 2007). The magnitude of ROLE is similar to the other aggregate measures included in the regressions. This indicates that the student–teacher gender interaction effect is as important for determining educational outcomes as the other factors. However, even if the size of our ROLE estimate is similar to the other aggregate measures that are commonly used in previous studies (e.g. Hanushek and Kimko, 2000; Barro and Lee, 2001), it may not be considered a major determinant for educational outcome. Nevertheless, despite the positive sign of ROLE that contradicts a gender role model effect, the significant results still imply that there is a gender interaction effect.

Given that ROLE captures a gender interaction effect only, the positive effect of ROLE can have several interpretations. Firstly, since ROLE measures the absolute difference between the fraction of female teachers and female students, the positive effect could simply imply that a gender imbalance is favourable regardless of the gender. In other words, being assigned an opposite gender teacher has a positive effect on the outcome. This can imply that same gender is not always optimal for enhancing the outcome, which is in line with previous studies claiming that male and female teachers are differently suited for teaching different subjects (Dee, 2007) Secondly, since there is little variation in the fraction of female students measure the largest variation in ROLE is reflected by the variation in FEMT. Thus, the positive effect might as well reflect a positive female role model effect for both boys and girls. It might be so those female teachers biologically are better suited to take care of and teach children in primary school. This implies that the higher fraction of female teachers in a country would enhance the educational outcome. However, this does not necessarily imply that male teachers cannot constitute role models but the influence male role models might have do not seem to enhance the educational outcome. Thirdly, the reason for not getting a negative effect of ROLE on educational outcome might be caused by the fact that we look at math and science scores, where boys are generally performing better than their female cohorts (Dee, 2007). As our dependent variable measures cognitive skills rather than curriculum based skills, the role model effect may not be as strong as it would have been if using a curriculum-based test score instead. Furthermore, the fact that female teachers can enhance girls' performances in math could further increase the proposed effect of ROLE in a math and science context. This is an indication that the gender role model effect differs across subjects, as seen in previous studies as well (e.g. Dee, 2007; Antecol, 2012).

It may be so that role model theory alone cannot explain the interaction effect. As previous studies have claimed, there might exist Pygmalion effects and negative stereotype threats as alternative mechanisms to the role model effect, that could further explain our findings. In other words, teachers' expectations and perception of their students in terms of stereotypes might induce a certain type of behaviour affecting educational outcomes (Antecol, 2012). Finally, in line with the reasoning of those who criticise the role model theory, it is inevitably possible that students do not wish to, or can, identify with their teacher no matter their gender. Furthermore, all teachers are, regardless of gender, not necessarily good role models (Driessen, 2007).

6.3 The effect of control variables

Looking at the results for the socio-economic factors we find that the education of adults has, as expected, a positive impact on educational outcome, and has a stable significant positive effect regardless of the composition of the control variables. This implies that the perceived importance of education has an effect on the educational outcome. GDP per capita has a positive effect on educational outcome when not controlling for FE and negative when controlling for FE. This finding is in line with the mixed results from previous research (e.g. Barro and Lee, 2001). A possible explanation for the negative effect when using FE, is that when controlling for cultural factors that possibly capture how efficient the educational system is, the level of income becomes less important in explaining differences in educational outcome across countries. Intuitively, a higher GDP might as well reflect higher income dispersion. In such a case, a high GDP can imply a less stable social environment, which may have negative effects on the learning environment and consequently educational outcome.

Turning to the school resource factors, we find a negative, although insignificant, effect of expenditure on educational outcome. The findings for both GDP and the expenditure on school are in line with those claiming that the effectiveness and interventions taken by the policy-making organs are more important for explaining differences in educational outcome than the amount of income or pure expenditure (Hanushek and Kimko, 2000). This is further confirmed, as expenditure is only significant in the regression without FE, as this regression does not capture these institutions. There is no way to assess the effectiveness of the use of resources, such as the quality of textbooks and facilities. A government can have high levels of school expenditures but if these resources are not used in the right way the expected positive effect on schooling outcomes will not occur.

The pupil-teacher ratio is never significant and has the incorrect sign compared to what we expected. This effect contradicts the theories saying that the quality becomes better when there are fewer students per teacher (Barro and Lee, 2001). Compared to Barro and Lee (2001) who use a dummy for East Asian countries instead of using region FE, our positive coefficient for PTR could reflect the different classroom effects of East Asian countries that are suggested to have a positive effect on educational outcome (Hanushek and Kimko, 2000). One underlying reason for this contradictory finding might imply that quality rather than quantity is what determines the educational outcomes. A low PTR ratio is good since teachers then get more time per student, but imposing policies to lower the ratio without considering changing the pedagogy or the quality of those who teach, or taking it even further, changing the mentality of parents and students might not lead to higher outcomes. Moreover, it could result from a case of over-controlling. A low fraction of expenditure could imply that there are no resources available, which inevitably leads to higher PTR. In other words, the effect of PTR might as well be captured already in the expenditure estimate. HOURS also contradicts our expectations, as more hours seem to have a negative impact on educational outcome. This effect however is very small and has little practical significance. This could also be an effect of over controlling as we have other resource measures, but when dropping HOURS some explanatory power is lost and the other resource estimates are not substantially altered. The reasons that this measure contradicts theoretical suggestions might be that it does not matter how many hours you spend in the classroom if what is taught is not pedagogical or relevant for your future and cognitive ability (Barro and Lee, 2001). This reasoning is similar to the reasoning of expected school length in previous studies. The coefficient for PRIVATE is negative rather than positive, which is expected when schools are given more autonomy in the decision making process. This result could intuitively reflect that private schools might be subject for harder controls and evaluations in order to be able to carry educational activities. Controlling for the fraction of private schools does not make us gain any significance in our regressions. This might arise as an effect of our data being on both a pure public level, and a combined public and private level. Another way of looking at this institutional variable is to conclude that it only shows how many schools are privately governed. Turning to the debate on gains in private welfare companies could possibly provide useful insights of the institutional measure. In some countries private companies are allowed to make profit on educational services, while they are being prohibited in others (Sveriges Radio, 2013). A measure of this might be better suited in explaining the impact of different governing systems.

6.4 Robustness

Undertaking our robustness regressions we get support for our model, as these results are similar to our main regression and findings of other studies. The most contradictory result is found when using European countries only. Generally, the robustness regressions show that ROLE appears to have a positive impact. However, it is only significant when including the East Asia dummy.

When estimating the same model with average data from primary and lower secondary school, all significance for ROLE is lost. On the one hand, this finding is consistent with previous research stating that lower secondary resource measures are not important to include when explaining differences in educational outcome using international test scores. On the other hand, it may imply that a role model effect is not supported when taking the higher level of education into account. This further confirms previous research that the role model effect has the largest effect in primary schooling. Using a subsample of European countries the estimates maintain their respective signs and significance. However, the significance for ROLE is lost. In other words, the base regression holds, but our main regression does not. The effect of ROLE decreases and comes closer to zero, and the standard deviation now can make the effect negative. One reason for the distorted results is that the countries and regions in Europe are more similar and an actual gender role model effect might not have the same impact on educational outcome as all are given equal opportunities in school. Thus we get an estimate closer to the expected effect as it does not capture other differences in educational outcome to the same extent as in the regressions with the full OECD sample. In the Europe sample regression, salaries have the greatest impact on educational outcome. When using the full sample, but with a dummy for East Asian countries, ROLE is still significant and of the same size as before. It remains robust. GDP gets its expected sign, implying that East Asia specific cultural factors affecting educational outcome is captured by GDP and thus when controlling for East Asia GDP receives its proposed positive impact on educational outcome. The evidence for GDP is mixed also in our study. We therefore realise that a shortcoming of our study is that we have not controlled for the income distribution. Such a measure could have been including the Gini index.[22] One possible explanation for the negative impact of GDP on educational outcome is that higher levels of GDP per capita induce higher income dispersion, meaning that country level inequality are bad for the educational outcome. This income dispersion is likely to have changed, at least slightly, over the 12-year time period due then is not constant over time and thus not captured by our fixed effect regression. Furthermore, in the regression with the East Asian dummy, PTR becomes significantly negative for the first time. This is in line with previous research (e.g. Hanushek and Kimko, 2000; Barro and Lee, 2001) stating that the East Asian countries have a different classroom mentality than their OECD cohorts and thus the effect of larger class sizes have a negative effect on educational outcome when controlling for East Asia. This finding supports the relevance of our model.

Controlling for FEMT and FEMS instead of ROLE, the effect of the other estimates changes slightly. The estimate for FEMT has a similar effect as ROLE has in the main regression, holding FEMS fixed. The net effect of FEMT and FEMS is different from the effect of ROLE implying that there is unobserved effects of teacher student interaction when not controlling for the difference in distribution between female teachers and students. The effect of FEMT and FEMS are positive. However, FEMS is not significant, which means that the female students solely do not explain differences in educational outcome.

[22] Gini Index: Summary measure of income inequality. "The Gini Index varies from 0 to 1, 0 indicating perfect equality where there is a proportional distribution of income. A 1 indicates perfect inequality where one person has all the income and no one else has any" (Noss, 2010, p. 1).

7 Conclusion and implications

The feminisation of primary school, implying that the number of male teachers is decreasing, is blamed for being one reason to why educational outcomes in some countries have worsened over the years. The purpose of this paper is to investigate whether this debate is motivated, i.e. whether there is a need for policy-makers to prioritise interventions that can increase the number of male teachers in the teacher force. Given the limited studies on student–teacher interaction in a broader, internationally comparable, context as well as with the current debate in mind, we ask ourselves whether there is a need for more male teachers in primary school? This paper makes an attempt to contribute to the studies on educational production and student–teacher interaction by investigating whether a gender role model effect is supported on the aggregate level and if it can help explaining differences in educational outcome across countries. Gender role model theory suggests that a student who is assigned a same-gender teacher performs better due to a role model effect. We create a measure of the absolute difference between the fraction of female teachers and the fraction of female students to reflect the suggested role model effect on the aggregate level. In an attempt to receive a causal effect of gender role models, we estimate an educational production function controlling for region and year fixed effect as well as commonly used aggregate socio-economic factors and school resources. The study uses the average of math and science scores from all five PISA surveys that have been conducted, constituting a time period of 12 years, as a proxy for educational outcome.

To answer our research question we state two hypotheses. Firstly, we ask whether there is a significant student–teacher gender interaction effect that can explain differences in educational outcome across countries. Our results strongly support this hypothesis. Also, we get an indication of practical significance of this effect since the estimated effect of the student–teacher gender interaction variable and the estimated effects of the other resource variables on educational outcome are similar in size. Secondly, we ask ourselves whether this student–teacher gender interaction effect has a negative sign, thus indicating that a gender role model effect prevails. We find a positive effect of a student–teacher gender imbalance on educational outcome. In other words, we find support that contradicts our second hypothesis, which states that a student–teacher gender imbalance has a negative effect on educational outcome. Our results indicate that a large imbalance between the fractions of female teachers and female students is positive for educational outcome. Since we do not find support for a gender role model effect, i.e. we see a positive rather than a negative effect of a student–teacher gender imbalance on educational outcome, we do not find any evidence of a need for more male teachers in primary school. We do not wish our findings to be directly interpreted as that there is no gain from increasing the number of male teachers to enhance educational outcome. Since this is a cross-country study using aggregate data the results have to be interpreted with caution. Also, we make a strong assumption that the fraction of female teachers and female students are representative for all classes in every country. However, we do find support on the aggregate level that there is an effect of student–teacher gender interaction and that this issue is motivated to consider in policy-making regarding the recruitment of teachers in the primary levels of schooling. We also get an indication that a gender role model effect in line with our second hypothesis may prevail if controlling for the fraction of female teachers, i.e. if controlling for the different characteristics for female and male teachers.

For further research, there is reason to evaluate and improve the measures of teacher background since these are likely to affect both input variables and educational outcome. The lack of data on less advanced countries force us to limit the study to OECD countries. As a result, the variation on the fraction of female students is unfortunately trivial. Using a larger country sample including less developed countries possibly provide a larger variation in this variable, and it would thus be interesting to investigate whether the gender role model effect would prevail in such a sample. Moreover, there might be a need for controlling for income dispersion to capture social inequalities affecting the educational outcome in a country. Race and ethnicity are other possible interaction effects that would be interesting to control for and study further as data becomes available. Moreover, as more data on non-cognitive skills becomes available for international comparison it would be interesting and favourable to include such data in the analysis as well.

Even though this study does not find a gender role model effect that can support a need for more male teachers in school, our results still indicate that the interaction between teachers and students, gender wise, matters for educational outcome. Hence, it is motivated to further study this interaction effect on both international and national levels, preferably using disaggregated data.

8 References

Alfonso, A. and St. Aubyn, M. (2006) 'Cross-country efficiency of secondary education provision: A semi-parametric analysis with non-discretionary inputs', *Economic Modelling*. 23 (3), pp. 476–491.

Antecol, H., Eren, O. and Ozbeklik, S. (2012) 'The effect of teacher gender on student achievement in primary school: Evidence from a randomized experiment', *IZA Discussion Paper,* No. 6453, Bonn: Institute for the Study of Labor.

Barro, R.J. (1991) 'Economic growth in a cross section of countries', *Quarterly Journal of Economics,* 106(2), pp. 407-443.

Barro, R.J. and Lee, J-W. (2001) 'Schooling quality in a cross-section of countries', *Economica.* 68 (272), pp. 465-488.

Bettinger, E.P. and Long, B.T. (2005) 'Do faculty serve as role models? The impact of instructor gender on female students', *American Economic Review,* 95(2), pp. 152-157.

Bishop, J. and Woessmann, L. (2004) 'Institutional effects in a simple model of educational production', *Education Economics* 12(1), pp. 17-38.

Carell, S.E. and West, J.E. (2010) 'Does professor quality matter? Evidence from random assignment of students to professors', *Journal of Political Economics,* 118(3), pp. 409-430.

Coleman, J.S., Campbell., E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D. and Robert L. York. (1966) *Equality of Educational Opportunity*, Washington DC: National Center for Educational Statistics.

Dee, T. (2007) 'Teachers and the gender gaps in student achievement', *The Journal of Human Resources,* 42(3), pp. 528-554.

Dolton, P. and Marcenaro-Gutierrez, O.D. (2010) 'If you pay peanuts do you get monkeys? A cross-country analysis of teacher pay and pupil performance', *Economic Policy*, 26(65), pp. 5–55.

Driessen, G. (2007) 'The feminization of primary education: Effects of teachers' sex on pupil achievement, attitudes and behaviour', *International Review of Education,* 53(2), pp. 183-203.

Ehrenberg, R. G., Goldhaber, D.D. and Brewer, D.J. (1995) 'Do teachers' race, gender, and ethnicity matter? Evidence from the national education longitudinal study of 1988', *Industrial and Labor Relations Review,* 48(3)*,* pp. 547-561.

Evans, M.O. (1992) 'An estimate of race and gender role-model effects in teaching high school', *Journal of Economic Education,* 23(3), pp. 209-217.

Haveman, R. and Wolfe, B. (1995) 'The determinants of children's attainments: A review of methods and findings', *Journal of Economic Literature,* 33(4), pp. 1829-1878.

Hanushek, E. A. (Summer 1979) 'Conceptual and empirical issues in the estimation of educational production functions', *Journal of Human Resources,* 14(3), pp. 351-388.

Hanushek, E.A. (1986a) 'The economics of schooling: Production and efficiency in public schools', *Journal of Economics Literature,* 24(7), pp. 1141-1177.

Hanushek, E.A. and Kimko, D.D. (2000) 'Schooling, labor-force quality, and the growth of nations', *American Economic Review,* 90(5), pp. 1184-1208.

Hanushek, E.A. (2002) 'Publicly provided education', *NBER Discussion Paper,* No. 8799, Cambridge: National Bureau of Economic Research.

Hanushek, E.A. and Woessmann, L. (2007) 'The role of school improvement in economic development', *NBER Discussion Paper*, No. 12832, Cambridge: National Bureau of Economic Research.

Hanushek, E.A. and Woessmann, L. (2009) 'Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation', *NBER Discussion Paper,* No. 14633, Cambridge: National Bureau of Economic Research

Hanushek, E.A., Link, S. and Woessmann, L. (2011) 'Does school autonomy make sense everywhere? Panel estimates from PISA', *IZA Discussion Paper,* No. 6185, Bonn: Institute for the Study of Labor.

Hanushek, E.A. and Woessmann, L. (2011) 'The economics of international differences in educational achievement', In: Handbook of the Economics of Education, Vol. 3, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. Amsterdam: North Holland: pp. 89-200.

Hanushek, E.A. and Woessmann, L., Schivardi, F., and Pistaferri, L. (2011) 'How much do educational outcomes matter in OECD countries? ', *Economic Policy,* 26(67) pp. 429-491.

Hoxby, C. (2000) ‘Peer effects in the classroom: Learning from gender and race variation’, *NBER Discussion Paper*, No. 7867, Cambridge: National Bureau of Economic Research.

Jurges, H. and Schneider, K. (2004) 'International differences in student achievement: An economic perspective', *German Economic Review*, 5(3), pp. 357–380.

Lucas Jr., R.E. (1988) 'On the mechanisms of economic development', *Journal of Monetary Economics,* 22, pp. 3-42.

Mankiw, N.G., Romer, D. and Weil, D.N. (1992) 'A contribution to the empirics of economic growth', *Quarterly Journal of Economics,* 107(2), pp. 407-437.

Nixon, L.A. and Robinson, M.D. (1999) 'The educational attainment of young women: Role model effects of female high school faculty', *Demography,* 36(2), pp. 185-194.

Noss, A. (2010) *Household Income for States: 2008 and 2009*, American Community Survey Briefs: US Census Bureau.

OECD (2002-2013), *Education at a Glance: OECD Indicators – 2002-2013,* Paris: OECD Publishing.

OECD (2005-2013) *OECD Factbook 2005-2013: Economic, Environmental and Social Statistics*, Paris: OECD Publishing.

OECD (2014) *PISA 2012 Results: What Students Know and Can Do - Student Performance in Mathematics, Reading and Science (*Volume I), Revised edn, Paris: PISA, OECD Publishing.

Piopiunik, M., Schwerdt, G. and Woessmann, L. (2013) 'Central school exit exams and labor-market outcomes', *European Journal Of Political Economy*, 31, pp. 93-108.

Romer, P. (1990a), 'Endogenous technological change', *Journal of Political Economy,* 98(5), pp. 71-102.

Romer, P. (1990b) 'Human capital and growth: Theory and evidence', *Carnegie-Rochester Conference Series on Public Policy*, 32, pp. 251-286.

Sveriges Radio (11 May 2013) *Solberg: Fler friskolor i Norge – men utan vinst,* Available at: *http://sverigesradio.se/sida/artikel.aspx?programid=83&artikel=5530169* [accessed 14 May 2014].

UNESCO Institute for Statistics (2014) *ISCED: International Standard Classification of education,* Available at: *http://www.uis.unesco.org/Education/Pages/international-standard-classification-of-education.aspx* [accessed 14 May 2014].

United Nations Statistics Division (2013) *Composition of macro geographical (continental) regions, geographical sub-regions, and selected economic and other groupings,* Available at: *https://unstats.un.org/unsd/methods/m49/m49regin.htm* [accessed 14 May 2014].

Wooldridge, J (2013) *Introductory Econometrics*, 5th edn., Andover: Cengage Learning, Inc.

Woessmann, L. (2003b) 'Schooling resources, educational institutions and student performance: the international evidence', *Oxford Bulletin of Economics & Statistics* 65, pp. 117- 170.

Woessmann, L., (2007a). 'International evidence on expenditure and class size: A review', *Brookings Papers On Education Policy 2006/2007*, pp. 245-272.

9 Appendix

TABLE A.1: VARIABLE SPECIFICATION AND DATA SOURCES

Name	Label	Definition	Source
PISA	PISA Score	Performance of 15 year-olds on the PISA mathematics and scientific literacy scale	2000-2009: OECD Factbook 2005-2012 2012: PISA 2012 Results: What Students Know and Can Do (Volume I) Table I.A
LOG_GDP	GDP per Capita	Gross domestic product per Capita (expenditure approach) in US $, current prices, current PPPs	UNESCO Institute for Statistics, Dataset 1. GDP
EDUC	Adult attainment	Attained a tertiary education degree (ISCED 5 or 6), 25-64 year-olds (%)	2000: OECD, stat, Dataset: Education at a Glance, Chapter A 2003: OECD Factbook 2009 2006: OECD.Stat, Dataset: Education at a Glance, Chapter A 2009: OECD.Stat, Dataset: Education at a Glance, Chapter A 2012: UNESCO Institute for Statistics, dataset Education
GOVEXP	Government Expenditure, per student	Government Expenditure per Student as % of GDP per Capita (%)	UNESCO Institute for Statistics, Dataset: Education
LOG_SAL		Annual statutory teachers' salaries in public institutions at starting salary, by level of education, in equivalent US dollars converted using PPPs	2000: Education at a Glance 2002, table D6.1 2003-2012: Education at a Glance 2005-2013, table D3.1
PTR	Pupil-teacher ratio	Average number of students per teacher at a specific level of education in a given school year	UNESCO Institute for Statistics, Dataset: Education
HOURS	Teaching Hours per Year	Net statutory contact time in public institutions, in hours per year, by level of education	2000: Education at a Glance 2013, table D4.2 2003: Education at a Glance 2005, table D4.2 2006-2012: Education at a Glance 2013, table D4.2
POPG	Population growth	Population growth rates (annual)	OECD.Stat, Dataset: ALFS Summary Table
FOREIGN	Foreign Population (%)	Stocks of foreign population by nationality in OECD countries as % of total population (%)	Key Statistics on migration in OECD Countries, table A.5
FEMT	Fraction of female teachers	Percentage of teachers who are female, by level of education (%)	UNESCO Institute for Statistics, Dataset: Education
FEMS	Fraction of female students	Percentage of students who are female, by level of education (%)	UNESCO Institute for Statistics, Dataset: Education
GDP deflator	GDP Deflator	GDP deflator (base year varies by country)	UNESCO Institute for Statistics, Dataset: Demographic and Socio-Economic
CPI	Consumer Price Index	Consumer Prices (All items)	OECD.Stat, Dataset: Consumer Prices (MEI)

TABLE A.2: PISA PARTICIPATION AND SCORES

	2000	2003	2006	2009	2012
NORTHERN EUROPE					
Denmark	497.50	494.75	504.46	501.31	499.25
Estonia	-	-	-	519.97	530.98
Finland	537.00	546.25	555.84	547.29	532.10
Iceland	505.00	504.90	498.17	501.13	485.48
Ireland	508.00	504.10	504.9	497.56	511.75
Norway	499.50	489.70	488.19	498.92	491.95
Sweden	511.00	507.55	502.85	494.67	481.53
United Kingdom	530.5	-	505.11	503.06	504.03
WESTERN EUROPE					
Austria	517.00	498.30	508.16	495.12	505.66
Belgium	508.00	519.05	515.36	510.92	509.99
France	508.50	511.00	495.38	497.50	496.98
Germany	488.50	502.65	509.72	516.59	518.82
Luxembourg	444.50	488.00	488.16	486.50	490.53
The Netherlands	-	536.10	527.76	524.03	522.51
Switzerland	512.50	519.80	520.59	525.26	523.11
SOUTHERN EUROPE					
Greece	454.00	462.95	466.29	468.11	459.85
Italy	467.50	476.05	468.54	485.87	489.43
Portugal	456.50	466.85	470.23	489.92	488.17
Slovenia	-	-	511.64	506.62	507.63
Spain	483.50	486.10	484.00	485.87	490.38
EASTERN EUROPE					
Czech Republic	504.50	519.90	511.36	496.66	-
Hungary	492.00	496.65	497.43	496.41	485.67
Poland	476.50	494.00	496.62	501.44	521.66
Slovak Republic	-	496.55	490.27	493.47	476.42
WESTERN ASIA					
Israel	433.50	-	-	450.86	468.28
EASTERN ASIA					
Japan	553.50	540.85	527.25	534.21	541.57
Korea	549.50	540.35	534.80	542.11	545.78
OCEANIA					
Australia	530.50	524.70	524.70	520.81	512.82
New Zealand	532.50	522.20	526.19	525.65	507.69
NORTHERN AMERICA					
Canada	531.00	525.60	530.74	527.76	521.77
United States	496.00	487.10	481.63	494.70	489.39

SOUTHERN AMERICA					
Chile	399.50	-	-	434.26	433.78
Mexico	404.50	395.05	407.65	417.21	414.10

Notes: "-" denotes non-participating Southern America consists of both Central and South American countries. The scores presented is the average of the country means in Math and Science tests

TABLE A.3: CORRELATION MATRIX

Variable	LOG_GDP	EDUC	EXP	LOG_SAL	PTR	HOURS	ROLE	FEMT	FEMS	POPG	PRIVATE
LOG_GDP	1.0000										
EDUC	0.4673	1.0000									
EXP	-0.0128	0.0672	1.0000								
LOG_SAL	0.8065	0.4504	-0.0155	1.0000							
PTR	-0.1933	0.1894	0.4206	0.0244	1.0000						
HOURS	0.3601	0.1947	-0.3290	0.4029	0.2419	1.0000					
ROLE	-0.2553	-0.3843	0.2533	-0.4490	-0.2046	0.0153	1.0000				
FEMT	-0.2339	-0.3674	0.2558	-0.4443	-0.2224	0.0131	0.9989	1.0000			
FEMS	0.4508	0.3572	0.0527	0.1007	-0.3741	-0.0458	-0.0283	0.0193	1.0000		
POPG	0.5202	0.2852	-0.1631	0.4902	0.1022	0.3676	-0.3236	-0.3236	0.1183	1.0000	
PRIVATE	.01329	0.1243	-0.0277	0.1863	-0.2731	0.0925	-0.02007	-0.2007	0.1318	0.1095	1.0000

TABLE A.4: BASE REGRESSIONS

Independent variables	*Dependent variable:* PISA		
	B1	B2	B3
LOG_GDP	34.1334**	31.1480**	-23.6464***
	(13.6411)	(12.2446)	(7.9371)
EDUC	0.8923**	0.6557*	0.7798***
	(0.3866)	(0.3945)	(0.2786)
EXP	-0.0665	-1.0304	-0.3639
	(0.8519)	(0.8816)	(0.6810)
LOG_SAL	-2.0023	-7.3837	18.3770**
	(7.7696)	(7.3313)	§
PTR	0.7217	-1.4907	0.7045
	(1.0047)	(0.9746)	(0.7427)
HOURS	-0.0952***	-0.0589**	-0.0216
	(0.0208)	(0.0250)	(0.0263)
EAST ASIA	-	57.3747***	-
		(14.6239)	
CONSTANT	216.1685**	325.8441***	561.4621***
	(107.3139)	(92.7642)	(87.9703)
FE	-	-	Yes
No. of obs.	91	91	91
R^2	0.3448	0.4597	0.8603
Adjusted R^2	0.2980	0.4141	0.8254

*Notes: Robust standard errors in parentheses. ***Significant at 1 % level (p<0.01)**Significant at 5 % level (p<0.05), *Significant at 10 % (p<0.1)*

TABLE 5.1 ROLE

	2000	2003	2006	2009	2012
Australia	-	-	-	-	-
Austria	40.10	41.56	40.73	41.17	42.30
Belgium	29.66	29.04	30.51	31.71	32.15
Canada	19.24	-	-	-	-
Chile	29.10	28.63	30.17	29.96	29.90
Czech Republic	35.85	35.75	46.24	49.13	48.53
Denmark	15.37	15.29	-	-	-
Estonia	37.85	38.27	41.40	45.25	44.66
Finland	22.91	25.97	27.02	29.71	30.12
France	31.26	32.04	33.27	34.07	33.98
Germany	32.65	34.23	35.34	36.34	35.83
Greece	8.41	13.88	15.62	16.79	-
Hungary	36.49	35.77	47.73	47.66	47.34
Iceland	-	-	-	31.56	31.36
Ireland	36.26	34.90	36.18	35.85	36.20
Israel	36.12	35.83	35.54	35.27	36.53
Italy	46.15	46.88	47.36	46.91	-
Japan	21.61	16.21	16.08	-	-
Korea	23.31	26.02	28.52	29.78	30.46
Luxembourg	17.39	20.54	22.71	22.70	25.02
Mexico	14.15	14.64	17.79	17.69	17.97
The Netherlands	-	-	-	-	-
New Zealand	35.22	35.46	34.83	35.00	34.46
Norway	-	-	-	-	-
Poland	35.06	36.16	35.71	35.25	36.53
Portugal	34.34	34.55	32.98	31.65	31.49
Slovak Republic	41.80	43.64	40.91	40.63	40.79
Slovenia	47.27	47.99	49.06	49.14	48.74
Spain	20.69	22.16	22.22	25.89	27.39
Sweden	31.06	31.07	32.37	33.02	33.45
Switzerland	-	-	-	-	-
United Kingdom	32.31	32.60	32.42	32.02	38.21
United States	38.08	39.31	39.58	37.72	38.24

Note: This table shows the difference between the fraction of female teachers (%) and the fraction of female. "-" denotes a missing value.

TABLE 5.2 ROLE

	2000	2003	2006	2009	2012
Min.	8.41	13.88	15.62	16.79	17.97
Mean	30.35	31.42	33.55	34.69	35.49
Max.	47.27	47.99	49.06	49.14	48.74

Note: This table shows the sample minimum, mean and maximum values of the ROLE measure.

TABLE 6: FEMS, FEMT AND ROLE

	Min.	Mean	Max.
FEMS	47.51	48.58	49.36
FEMT	56.83	81.96	97.63
ROLE	8.41	33.39	49.14