Essays on Human Capital

Balint Lenkei

M00274872

Director of Studies: Dr. Michela Vecchi
Advisors: Dr. Ericka G. Rascón Ramírez and Dr. Quang Nguyen

Middlesex University London

This thesis is submitted to Middlesex University in partial fulfilment of the requirements for the Degree of Doctor of Philosophy

Date: 06/02/2020

Abstract

This thesis presents the results of my doctoral studies at Middlesex University. It contains three studies on human capital investment, education and cognitive abilities. The first study follows a macro approach and assesses the impact of various human capital measures on productivity. The main finding of this chapter is that human capital quality has a larger impact on economic growth than human capital stock, while the distribution of human capital also plays a role. The second chapter explores specific microeconomic issues of human capital by studying parental expectations in rural Peru. The main finding is that both time and risk preferences are important factors associated with expectations. Lastly, the third chapter deals with cognitive abilities by presenting the results of a meta-study on the Cognitive Reflection Test. The results suggest that males are likely to perform better on the test, while the monetary incentive have no significant impact on test performance. Furthermore, regarding the implementation of the test we find that overall students perform better, while having the test after an experimental session negatively impacts test performance.

Declaration

I, Balint Lenkei, declare that this thesis and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Table of contents

Introduction

This thesis presents the findings of my doctoral studies at Middlesex University titled 'Essays on Human Capital'. It consists of three main chapters dealing with the related topics of human capital, education and cognitive abilities. I am following a top down approach by first discussing macroeconomic issues related to human capital and later delving into two specific micro level applications including mother's educational expectations in developing countries and drivers of performance on a cognitive ability test.

In Chapter One *'Education, education inequality and growth'* I investigate the channels through which human capital affects economic growth by looking at both the direct impact of the investments in education and the way educational resources are distributed within society.

Investments in human capital have been long identified as important determinant of growth in macroeconomic theory (Becker 1964, Schultz 1981, Romer 1986). Education increases the human capital available in the labour force, which in turn enhances labour productivity, eventually leading to a higher equilibrium level of output (Mankiw et al. 1992). Education also increases the innovative capacity of an economy, and the newly obtained knowledge on new products, processes and technology stimulates growth (Lucas 1988; Romer 1990a). Furthermore, education may enable the transmission and diffusion of knowledge required to process and understand new information and to successfully implement new technologies developed by others, which in turn promotes economic growth (Benhabib and Spiegel 2005).

Despite the indisputable theoretical impact of human capital, there is mixed empirical evidence in this area in the literature (see Pritchett 2001, 2006; Lenkei et al. 2018).

There are several reasons that may have led to the mixed conclusions in the past including both empirical and methodological issues. First, there are measurement issues. Typical proxies for human capital in the macroeconomic literature are schooling attainment measures (average years of schooling) such as the ones provided by Barro and Lee (2013), Cohen and Soto (2007) or the Penn World Table (Feenstra et al 2015). An issue with these measures is that they assume that an extra year of schooling provides the same increase in skills and knowledge regardless of the education system of the country. This has led to another branch of studies in the literature suggesting that accounting for the quality of education may be a more appropriate determinant (Hanushek and Kimko 2000, Hanushek and Wößmann 2007). One of the objectives of the first chapter is to shed light on some of the unresolved issues that still plague the applied macro literature.

The second issue relates to the distribution of human capital. When increasing mean education levels in the population, governments (especially in developing countries) face the question whether they should allocate resources to improve education levels of the already educated segments or to involve the wider society to provide basic education and alleviate illiteracy rates. The chapter also recognises that it is not just the accumulation of human capital that plays a role in determining growth but also its equal (or unequal) distribution. If the ability of a nation's population is normally distributed, then a skewed distribution of educational opportunities will lead to welfare losses. Thus, an equitable distribution of human capital constitutes a pre-requisite for substantial gains in productivity (Thomas et al. 2002). Nevertheless, the empirical findings are not conclusive in the literature regarding education inequality, thus the chapter will aim to bring some resolution to this issue.

The third issue on human capital concerns which framework to apply to model the true relationship between human capital and growth. Although the Neoclassical model is theoretically sound, it does not necessarily capture the dynamic long-run relationship between the variables, particularly when estimated for a large number of highly heterogeneous countries. Therefore, based on the works of Eberhardt and Teal (2013) and Lenkei et al. (2018), the chapter adopts an Error Correction Model (ECM) representation, which controls for both the long-run (accumulation effect) and short-run (growth effect) of human capital on growth. This approach also provides a more general way of specifying the role of human capital as it encompasses the other models in the literature (first difference specification; Benhabib and Spiegel (1994); Sunde and Vischer (2015)). Furthermore, it allows to take full advantage of the characteristics of the dataset such as long time dimension and availability of a panel of countries.

The results suggest that changes in average years of schooling between 2003 and 2015 do not significantly affect growth by using any of the educational attainment measures (BL, PWT, CS). However, initial levels of human capital do have an impact in a way that higher levels of initial average years of schooling in 2003 significantly contributes to higher growth over the observed time period. The results on human capital quality provide some evidence that quality (captured by PISA mathematics score) has a somewhat stronger impact as the magnitude of our quality coefficients are larger, although their effect is only weakly significant. Negative coefficients are observed for educational inequality, however they remain insignificant.

In contrast, the ECM framework (preferred model) predicts changes in human capital stock have played an important role in increasing productivity across countries. This is consistent across all three measures of human capital (BL, CS and PWT). In addition, human capital coefficients are larger a sub-sample of developing countries. This is consistent with previous findings in the human capital literature (e.g. Lenkei et al 2018). The dynamic panel data model also finds significant evidence that in the long-run a more equal distribution of education within societies (measured by educational GINI) leads to higher changes in income per worker. Overall, the empirical finding using the ECM framework support the view that in the long-run government policies should focus on allocating national resources to improve educational attainment and also distribute these investments equally across the population, as educational inequality has a negative impact on output growth.

In the second chapter the focus remains on human capital but in a microeconomic context. An individual's skills and knowledge obtained through education can be transformed into human capital. In turn, as established in the first chapter, human capital can be an important determinant in boosting growth in the long-run. This is particularly important for developing countries as it facilitates the technological catch-up process to developed countries. However, despite some recent advancement in providing basic education to children in developing countries, 57 million children are still said to be out of school and over 100 million lack basic literacy skills (UNESCO 2015). There are several factors that may prevent efficient investment in education in developing counties. First, education has a high opportunity cost as it is long-term commitment in the sense that years of investment are required before the returns can be realised (Becker 1964). Second, education is also considered to be a risky investment for households due to uncertain school qualities. Third, there are several issues related to parental preferences and decisions that driving

Parental expectations are also considered to have a substantial economic and social impact on the household. High expectations lead children to set higher standards for their education and to make greater demands on themselves from an early age which in turn results in high achievement, better attendance and more positive attitudes towards school (Boocock 1972, Astone and McLanahan, 1991, Peng and Wright 1994, Reynolds 1998, Davis-Kean 2005, Yamamoto and Holloway 2010). At the same time low parental educational expectations can be transmitted through generations and they may create a vicious circle and reinforce poverty. In consequence, low parental educational expectations may reduce attendance, academic achievement and successful labour market outcomes for children.

Due to the long-term commitment and riskiness of educational investments in developing counties, Chapter 2, investigates whether parent's willingness to take risk and their patience impacts their educational expectations.

Data from a field experiment in Peru is used to analyse the relationship between expectations on children education attainment and parental risk and time preferences among rural households. After controlling for a set of household, parental and child characteristics I find that the risk-aversion parameter, sigma, is significantly and positively related to expectations on schooling choices. There are three main explanations to these results. First, parents with high levels of risk aversion report higher expectations about their children's education as they likely perceive education for their children a less risky option. Therefore, parents consider high levels of education as a safe option. Second, if at an early child age mothers have low expectations and consequently low schooling investment there is a risk of losing out on future returns to human capital which can significantly affect the wealth of a low-income household. Third, all subjects were participants of a governmental conditional cash transfer (CCT) programme. A risk averse parent makes sure that their child has adequate primary school attendance and one of the spill-over effects of the CCT program is that it may increase future expectations beyond primary level. This may have important policy implications as it shows the effectiveness of conditional cash transfer programmes in improving human capital accumulation.

Moreover, a higher degree of impatience (higher discount rate) among mothers is significantly correlated with low expectations on children's educational choices. This suggests that parents with high discount rates are more impatient to wait for the returns to schooling and therefore they may tend to underinvest in human capital.

While past studies analysed the role of risk and time preferences of parents in determining parental investments both in developed countries (e.g. Germany, Wölfel and Heineck (2012)) and developing countries (e.g. Uganda, Tanaka and Yamano (2015)), to the best of the author's knowledge no study has estimated the association between risk and time preferences of parents and parental expectations for the children in low income countries. This study is also different from previous studies in this field in the sense that there are no observations of actual parental behaviour but predictions under uncertainty about future children outcomes.

The third chapter of the thesis also deals with the micro dimension but now looks at individuals and their cognitive abilities. As the first chapter discusses, human capital quality (i.e. cognitive skills of a population) may plays a more important role in driving growth compared to the quantity (schooling attainment) (Hanushek and Wössmann, 2007). Cognitive abilities can be regarded as an important dimension of human capital; along with education, health and non-cognitive abilities (Heckman and Rubinstein 2001; Banks and Mazzonna 2012), with most policy makers accepting the notion that they can be a key driver of schooling outcomes. Furthermore, they are also fundamental for decision making and play a crucial in the formulation and subsequent execution of saving and consumption plans (Banks and Oldfield 2007). As discussed in Chapter 1, there is ongoing debate about measuring of human capital. In particular, for cognitive abilities educational outcomes have been often used as a proxy given that these abilities are not easily observed. However, there are an increasing number of alternative measures in economics that are gaining popularity, one of which is the Cognitive Reflection Test (CRT). In chapter three I delve deeper in the discussion and analysis of the main drivers of performance on the CRT by using a meta-study approach.

The CRT was first proposed by Frederick (2005) and since then has been extensively used in the Experimental Economics and Psychology literature. Frederick proposed the test based on a dual-system theory (e.g. Epstein 1994; Sloman 1996; Stanovich and West 2000; Kahneman and Frederick 2002) made up of two cognitive processes: System 1, executed quickly without much reflection and System 2, more deliberate and requiring conscious thought and effort. The questions in the CRT have an immediate (intuitive) incorrect response (System 1). However, the correct response requires some deliberation, i.e. the activation of System 2.

Frederick (2005) found that individuals with high CRT scores are more patient and more willing to gamble in the domain of gains. He also provided evidence that the CRT scores are highly correlated with some other tests of analytic thinking (e.g. American College Testing [ACT], Scholastic Assessment Test [SAT], and Writing Proficiency Test [WPT]) and that males on average score higher on the test. Toplak et al. (2011) claim that the CRT can be viewed as a combination of cognitive capacity, disposition for judgement and decision making. They argue that the CRT captures important characteristics of rational thinking that are not measured in other intelligence tests.

The use of the Cognitive Reflection Test as a covariate to explain behaviour in Economics and Psychology experiments has significantly increased in the past few years. Experiments have shown its usefulness in predicting behaviour. However, little is known about if the test is gender biased, whether incentives matter or how different implementation procedures impact outcomes.

The chapter reports the results of a meta-study of 118 Cognitive Reflection Test studies comprising of 44,558 participants across 21 countries. Regarding the gender bias it is found that (i) males perform better in every CRT questions, (ii) females are more likely to answer none of the questions correctly, and (iii) males are more likely to answer all three questions correctly. This observation is important if one is interested in constructing samples based on cognitive ability. This could lead to strong (gender) sample imbalance. (For instance, if one uses three correct answers as a selection criterion then the sample is disproportionately biased towards males). The second finding is that there is no statistical evidence to support the argument that monetary incentives may play an important role in improving CRT performance. Albeit limited (no data on the amount, or how subjects were paid), this result is important as it shows that incentives may not be strongly relevant for the implementation of the CRT. Therefore, it may increase the effort that test takers exert when attempting to answer questions, but on average it does not necessarily result in higher CRT scores.

Regarding implementation procedures, taking the test at the end of an economic experiment negatively impacts performance. Conducting the test later decreases the probability of obtaining correct answers; meanwhile, the probability of obtaining none correctly is increased. This result is interesting as it points towards the fact that increased cognitive load could be an important determinant of performance in the CRT. When comparing student and non-student populations, we found that students are more likely to answer all three questions correctly compared to non-students, and less likely to have zero correct answers. This tells us that the explanatory power of the CRT may be affected by population differences. In addition, mixed evidence obtained on whether the sequence of questions matters. Finally, it is found that computerised tests marginally improve test results.

References

Astone, N.M., McLanahan, S. (1991) Family structure, parental practices and high school completion. American Sociological Review, 56(3): 309-320.

Banks, J., Mazzonna, F. (2012) The Effect of Education on Old Age Cognitive Abilities: Evidence from a Regression Discontinuity Design, Economic Journal, 122(560): 418-448.

Banks J., Oldfield Z. (2007) Understanding pensions: Cognitive functions, numerical ability and retirement saving. Fiscal Studies, 28(2): 143–170.

Benhabib, J., Spiegel, M.M., 1994. The role of human capital in economic development evidence from aggregate cross-country data. Journal of Monetary Economics, 34: 143–173.

Boocock, S.P. (1972) An Introduction to the Sociology of Learning. Boston: Houghton Mifflin Co.

Davis-Kean, P.E. (2005) The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. Journal of Family Psychology, 19: 294 –304.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. American Psychologist, 49, 709–724.

Frederick, S. (2005). Cognitive reflection and decision making. Journal of Economic Perspectives, 19(4), 25-42.

Hanushek, E.A., Wössmann, L. (2007) The role of education quality in economic growth. Policy Research Working Paper 4122. Washington, DC: World Bank.

Heckman, J.J., Rubinstein, Y. (2001). The Importance of Noncognitive Skills: Lessons from the GED Testing Program. American Economic Review, 91 (2): 145–149.

Wölfel, O., Heineck, G. (2012). Parental Risk Attitudes and Children's Secondary School Track Choice. Economics of Education Review, 31(5): 727–743.

Kahneman, D., Frederick, S. (2002) Representativeness revisited: At- tribute substitution in intuitive judgment. In T. Gilovich, D. Griffin and D. Kahneman (Eds.), Heuristics and biases: The psychology of intuitive judgment, 49-81, New York: Cambridge University Press.

Krueger, A., Lindahl, M., 2001. Education for growth: why and for whom? J. Econ. Lit. 39, 1101–1136.

Lenkei, B., Mustafa, G., Vecchi, M., (2018). Growth in emerging economies: is there a role for education? Economic Modelling, 73, 240-253.

Lucas, R.E., 1988. On the mechanics of economic development. J. Monetary Econ. 22, 3–42.

Mankiw, N.G., Romer, D., Weil, D.N., 1992. A contribution to the empirics of economic growth. Q. J. Econ. 107, 407–437.

Peng, S.S., Wright, D. (1994) Explanation of academic achievement in Asian American students. Journal of Educational Research, 87(6): 346–352.

Pritchett, L. (2001). Where has all the Education gone? World Bank Economic Review, 15: 367–391.

Pritchett, L. (2006) Does learning to add up add up? The returns to schooling in aggregate data. Handbook of the Economics of Education, 1: 635-695.

Reynolds, A.J. (1998) Resilience among black urban youth prevalence, intervention effects, and mechanisms of influence. American Journal of Orthopsychiatry, 68(1), 84-100.

Romer, P.M., 1990. Endogenous technological change. Journal of Political Economy, 98: 71–102.

Sloman, S.A. (1996) The empirical case for two systems of reasoning. Psychological Bulletin, 119, 3–22.

Stanovich, K.E., & West, R.F. (2000). Individual differences in reasoning: Implications for the rationality debate? Behavioral and Brain Sciences, 23, 645-665.

Tanaka, Y., Yamano, T. (2015) Risk and time Preference on schooling: Experimental evidence from a low-income country, GRIPS Discussion Paper 14-24.

Thomas, V., Wang, Y., Fan, X. (2002) A New Dataset on Inequality in Education: Gini and Theil Indices of Schooling for 140 Countries, 1960-2000, World Bank.

Topel R. (1999). Labor Markets and Economic Growth, in eds. Ashenfelter, O. and Card, D., Handbook of Labor Economics, Amsterdam: North Holland.

Toplak, M.E., West, R.F., & Stanovich, K.E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. Thinking & Reasoning, 20(2), 147-168.

Yamamoto, Y., Holloway, S.D. (2010) Parental expectations and children's academic performance in sociocultural context, Educational Psychology Review, 22: 189-214.

Chapter 1.

Education, education inequality and growth

This study investigates the channels through which human capital affects economic growth. We look at both the direct impact of investments in education and at the way educational resources are distributed within society. In fact, this study recognises that it is not just the accumulation of human capital to play a role in determining growth but also its equal (or unequal) distribution. Our analysis relies on a variety of techniques and measures of human capital and on a measure of educational inequality based on the Gini coefficient. First, using a long differences model we compare the performance of different measures of human capital, capturing both the quantity and quality of education, as well as education inequality. Second, we assess the long run relationship of human capital accumulation and educational inequality on growth using a method that encompasses previous specifications and estimates both short-run and long run effects. Overall, our results support the view that government policies should focus on improving the quality of education and allocating national resources to education equally across the eligible population, as educational inequality has a negative impact of output growth in the long-run.

1. Introduction

In the 21st century one of the most important policy issues facing governments is education and educational investments. Education may not only be beneficial for the individual but there might be spill-over effects at the macro level that would provide support for educational reforms in both developed and developing countries (Sianesi and Van Reenen 2003).

Education increases the human capital present in the labour force, which in turn enhances labour productivity, eventually leading to a higher equilibrium level of output (e.g. Mankiw et al. 1992). Education also increases the innovative capacity of an economy, and the newly obtained knowledge on new products, processes and technology stimulates growth (e.g. Lucas 1988; Romer 1990a). Furthermore, education may enable the transmission and diffusion of knowledge required to process and understand new information and to successfully implement new technologies developed by others, which in turn promotes economic growth (Benhabib and Spiegel 2005)[1].However, despite the widely shared view that human capital is an important determinant of growth in both macro- and microeconomic theory, there is

[1] At the individual level education has been found to contribute to better health outcomes (Cutler and Lleras-Muney 2008), lower fertility rates (Martin 1995, Basu 2002) and improvement in the health of women's children (Currie and Moretti 2003); it also reduces the incidence of criminal activity (Lochner and Moretti 2004) and helps in overcoming addictions (Sander 1995); furthermore, it also supports marriage market outcomes (Lefgren and McIntyre 2006) and more rational decision making (Goll and Rasheed 2005).

mixed empirical evidence in this area (For extensive review of the literature see Topel 1999; Krueger and Lindahl 2001; Pritchett 2001, 2006; Lenkei et al. 2018).

The objective of this study is to provide some resolution to the mixed evidence and unresolved issues on the relationship between human capital and growth that are still present in the macroeconomic literature. Firstly, it's still unclear whether the quantity or the quality of human capital plays the main role in accounting for growth. We attempt to contribute to the debate through contrasting a number of measures of both quantity and quality of human capital using the most up to date datasets. Our main human capital quantity proxies are average years of schooling measures including the Barro Lee, Penn World Table and Cohen and Soto datasets. Despite the imperfection of these variables (e.g. De la Fuente and Domenech 2006), their advantage is that they are available for a large number of countries for a long time period. In order to measure the quality of human capital we make use of the PISA test scores, which allow a standardised cross-country comparisons of students' cognitive abilities and the overall effectiveness of education systems.

Second, we shed light on whether a more equal distribution of human capital within the general population is associated with faster growth. Despite governmental efforts across the globe to reduce inequality in education over the past few decades, in many countries gaps in education between various groups is still substantial. If people's abilities are normally distributed, then a skewed distribution of education opportunities represents large welfare losses. Assuming that the amount of human capital of a person is the outcome of a combination of abilities and education, inequality in education would therefore lower the average level of human capital in the economy and consequently slow economic growth (Klasen 2002). However, the empirical literature provides mixed evidence on its impact, with some studies generally supporting the argument that inequality hinders growth (Castello and Domenech 2001; Thomas et al. 2002), while others do not find significant effects (Ram 1984; Foldvari and van Leeuwen 2011) or suggest that the impact of completed higher education is a more important driver of economic prosperity compared to general accessibility to lower levels of schooling (Gennaioli et al 2013; Squicciarini and Voigtlander 2015; Castello-Climent et al. 2017). Hence, the main issue is not only whether to devote Government resources to education, but whether such resources should be invested on a restricted elite or towards mass education. The objective of this study is to shed light on these issues in the applied macroeconomic literature.

The third contribution of this study is related to the applied analytical approaches. Besides making use of traditional methods of empirical analysis such as the first differences specification, the study investigates the role of human capital on growth using an innovative analytical framework. This method allows us to test directly the validity of the restrictions imposed by some of the most commonly used empirical models e.g. first differences model. Following the work of Eberhardt and Teal (2013) and Lenkei et al. (2018), we adopt an Error Correction Model (ECM) representation that controls for both the long-run (accumulation effect) and short-run (growth effect) of human capital on economic growth.

Over the past few decades economic and financial integration between countries is ever increasing. Thus, our panel data may exhibit cross-sectional dependence in the error terms due to common shocks such as the global financial crisis (Chudik et al. 2011) and the presence of spill-overs (Eberhardt et al. 2013). If the unobserved components that create interdependencies across-cross sections are correlated with the independent variables our estimates become biased and inconsistent (De Hoyos and Sarafidis 2006). Therefore, our econometric approach involves using the mean group estimator, while also controlling for heterogeneity and unobserved common factors causing cross-sectional dependency (Eberhardt and Teal, 2013; Eberhardt and Presbitero 2015).

Overall our result provide evidence that the quality of human capital is marginally more important in explaining growth than the quantity of human capital. In addition, we found that in the long run the increase in average years of schooling and the decrease in educational inequality positively contribute to productivity growth. Overall our results support the argument that instead of focusing on educating the elites, more resources should be allocated to the educational involvement of a larger share of the population.

The remaining of the study is structured as follows. In the next section we provide some background information on the literature in investments in human capital. Section 3 describes the analytical framework used for our analysis, while Section 4 provides a descriptive analysis of the data used in the study. Section 5 presents the results and discussion of the empirical analysis, while Section 5 concludes the study.

2. Background

There are several reasons that may have led to mixed conclusions on the role of human capital for economic growth in the past, which include both methodological and empirical issues. The first issue is how to measure and compare across countries and time the competencies and skills of individuals at the macro level. The challenge is to find a good proxy for the conceptual human capital investment in theoretical models.

Over the past few decades, the literature has been using a diverse range of measures (Mason et al. 2012). One of the most commonly used proxy is based on the average years of schooling for the population and the percentage of population receiving different levels of education from Barro and Lee (BL) (Barro and Lee, 2013). The main advantage of these measures is that they are available for several countries for extended time periods; however many researchers argue that they measure attendance rather than attainment and as such they are input rather than output variables (e.g. Sianesi and Van Reenen 2003; De la Fuente and Domenech 2006; Mason et al. 2012). Furthermore, they are based on formal education only, without accounting for other human capital effects such as job training or home education; they do not capture quality of education only the quantity; and that different educational levels have varying impact on growth (Lenkei et al. 2018). Cohen and Leker (2014) have developed an alternative dataset for average years of schooling based on Cohen and Soto (2007) (CS). The most important differences between the BL and CS datasets is that CS employs a different approach to extrapolate missing data, use more census observations and change values they consider implausible (Ciccone and Papaioannou 2009). The availability of the Barro-Lee (BL) and Cohen and Soto (CS) dataset provides an opportunity to compare both. Another broadly used educational variable for human capital is available in the Penn World Table (PWT 9.0) (Feenstra et al 2015). The PWT combines both the BL and CS measures of human capital into one dataset by correcting for their imperfections.

Studies based on these measures have produced mixed evidence on the impact of human capital on growth (for a discussion on data quality of human capital measures see de la Fuente and Domenech 2006, Cohen and Soto 2007, Portela et al 2010, Mason et al 2012).

Attainment of formal qualifications such as number of degrees have been also extensively used in the literature (e.g. Shapiro 2006), as they have the advantage to better capture the quality of human capital. However, these measures suffer from the issue of international comparability across diverse educational systems. Thus, another emerging segment of the human capital literature deals with the link between internationally comparable measures of education quality and the ability of countries to grow over time. This literature argues that the cognitive skills of a nation (as a measure of human capital quality) are more important in measuring the impact of human capital on growth compared to years of schooling (e.g. Hanushek and Kimko 2000; Hanushek and Woessmann 2007), because using the educational attainment measure assumes that an extra year of schooling provides the same increase in skills and knowledge regardless of the education system of the country. For instance, it is assumed an additional year of schooling in a developing country would create the same increase in productive human capital as in the US or EU. This ignores the cross-country differences in educational quality as an extra year of schooling delivers different skills depending on the efficiency of a particular education system to transfer knowledge to students. Moreover, many countries have reached a point where expanding the accessibility to basic education provides little marginal benefits for growth and instead governments should focus on improving the quality of education (Hanushek 2010).

Although we use the words schooling and human capital synonymously, an important distinction has to be made between schooling on one hand and skills and knowledge on the other. Therefore, another drawback of years of schooling as a measure of human capital is that it assumes that all human capital and skills is obtained through formal schooling (Hanushek and Woessman 2007). This ignores the possibility that there are various other factors outside the class room that contribute to improvement in cognitive skills and knowledge including family environment, friends, outside school activities, private tutors and on-the-job training. Thus, a measure of cognitive skills would be more suitable for cross-country comparisons of human capital. Recent research has identified these skills as an important dimension and there is very large payoff to them. Individuals with higher cognitive skills systematically do better than the ones with less, and nations with more skilled population tend to grow faster (Hanushek and Kim 1995, Hanushek and Kimko 2000). Some evidence suggests that these returns to education are even higher in emerging economies compared to developed countries (Knight and Sabot 1990; Behrman et al 2008), and may even increase over time due to spill-over effects (Lucas 1988). This allows developing countries with less technological capabilities to catch up more rapidly to technological leaders (Barro 1991).

Much of the debate about educational quality, which also captures cognitive skills, is centred around how to measure it. It is a challenge to find the same standardised measure that compares cognitive skills, and thus the quality of education, across countries and time with such diverse educational systems. In the literature the most commonly used proxies for quality are tests conducted by international agencies measuring students' cognitive skills, particularly focusing on science and maths. Among these international tests one popular measure is the PISA test score. For OECD countries the PISA scale historically has a mean of 500 and a standard deviation of 100, while developing countries that participate in the tests usually perform substantially lower. These measures are found to be positively and significantly related to economic growth in the past (Hanushek and Kimko 2000, Hanushek and Woessmann 2007). Furthermore, years of schooling measures such as BL tend to become insignificant when labour force quality measures are included, indicating that they are likely to be highly correlated.

Educational inequality

When increasing mean education levels in the population, country governments face a trade-off: should they focus on improving the education levels of the already educated segment of the society or instead they should allocate resources on involving the wider population to alleviate illiteracy rates and to provide basic education (e.g. Birdsall 1996)? The former would lead to larger, while latter to lower educational inequalities within nations. This dilemma has led to another area of human capital research that has been receiving increasing attention in recent years; the extent to which inequality in educational attainment within a society can hinder economic growth.

If skills are normally distributed within the population then a skewed distribution of educational opportunities will lead to substantial welfare losses. Education inequality represents an opportunity cost in the form of foregone salaried employment and earnings in an economy (Thomas et al. 2002). Thus, an equitable distribution of human capital constitutes a pre-requisite for improvement in productivity and reduction in poverty at the macro level. Education also builds assets and improves social welfare by its spill-over effects on health, political participation, reduction in crime rate, social trust (Lochner 2011; Hout 2012; Borgonovi and Burns 2015; Hooghe et al. 2015) which in turn will positively affect growth (Green and Henseke 2016). Therefore, ensuring access to educational opportunities is essential for the welfare of both developed and developing nations.

In the past few decades, both developing and developed countries have made a great effort to lower illiteracy and to involve the wider society into receiving education. It is estimated that inequality in the distribution of education has been more than halved as the average Gini coefficient dropped from 0.55 in 1960 to 0.28 in 2005 (Castello-Climent and Domenech 2014). Yet, the empirical findings in the literature are not always conclusive about the impact of education inequality. Some early studies use standard deviation of educational attainment in growth regressions to capture the effect human capital inequality. For example, using the standard deviation of years of education as a measure of educational dispersion for a sample of 43 countries Birdsall and Londono (1997) analyse the impact of initial distribution of land, income and human capital on output growth and obtain a negative effect for education inequality. In a similar manner, Lopez et al. (1998) find that for a panel of 12 Asian and Latin American countries between 1970 and 1995, the standard deviation of education tends to have a negative impact on per capita income. However, one issue with standard deviation is that it is an absolute measure of dispersion thus it does not control for differences in the mean of the distribution.

Using the educational Gini coefficient constructed from the Barro and Lee (1996) data, Thomas et al. (2002) and Castello and Domenech (2002) also find that a more equal distribution contributes to higher income per capita. Inequality in educational achievement and earnings inequality are also found to be correlated (Bedard and Ferral 2003, Blau and Khan 2005).

However, some other studies do not find similar significant effects of human capital inequality either through using variance of schooling (Ram 1984) or educational Gini coefficient (Foldvari and van Leeuwen 2011) measures. Furthermore, some other studies even find that higher inequality leads to higher productivity. For example, using a sample of 29 Indian states Gille (2015) obtains a negative relationship between equality in education and income per capita. While some other studies that look at the contribution of the upper tail of the human capital distribution, highlight that the impact of completed higher education is a more important driver of economic prosperity compared to general accessibility to lower levels of schooling (Castello-Climent and Mukhopadhyay 2013, Gennaioli et al. 2013, Cantoni and Yuchtman 2014, Squicciarini and Voigtlander 2015, Castello-Climent et al. 2017).

In addition, existing studies often use the classical first differences model to estimate the impact of inequality, which may not be able to capture these effects when using a large sample of heterogeneous countries (Islam 1995, Mason et al. 2012). Furthermore, the first differences model captures short run economic effects and does not account for long run relationships. Therefore, we also investigate whether a more equal distribution of human capital within the general population is associated with faster growth accounting for both short-run and long-run effects. We choose to utilise the Gini coefficient index (van Leeuwen and van Leeuwen-Li, 2015) in analysing educational inequality as it is normally the one also used in international comparisons of income distribution[2,3].

[2] There has been a great interest among researchers in assessing inequality through the use of such a simple index as the educational Gini coefficient. Despite its popularity there are some limitations associated with it. Deininger and Squire (1996) argue that changes in the coefficient do not explain whether educational redistribution is from the top educated to middle educated or an increase by the share of the bottom educated at the expense of the middle educated. Frankema and Bolt (2006) suggest that the lower the average years of schooling in a country, the bigger the effect of the gap in average years of education between individuals on the Gini coefficient. Therefore, a higher average level of educational attainment lowers the Gini coefficient.

[3] The issue with other measures such as the standard deviation is that it is an absolute measure of dispersion therefore it does not control for differences in the mean of the distribution.

3. Analytical Framework

The literature has also raised the issue of how to model the true relationship between human capital and growth. Therefore, in this section we provide a discussion on the empirical implementation of our analysis that involves various data sources and estimation techniques, allowing us to account for the different growth channels of human capital, check the robustness of our results and compare them with existing evidence.

3.1 Theoretical model

We begin our analysis between human capital and growth by utilising the Solow (1956) growth model. The three-factor aggregate production function with labour and capital (physical and human) as inputs can be represented as:

$$Y_{it}=A_{it}\,F_i(L_{it},\,K_{it},\,H_{it}) \tag{1}$$

where Y_{it} denotes the level of aggregate output in the economy, L_{it} is the quantity of labour (or number of workers or hours worked), K_{it} is an estimate of the capital services provided by the stock of equipment and structures used to produce goods and services and H_{it} is the education and skill level of a country's workers. A_{it} is a multiplicative factor referred to as total factor productivity and it reflects the general level of productivity or technology in the economy.

Assuming a Cobb-Douglas production function and constant returns to scale, we can modify the Equation (1) by dividing it by L_{it} which gives us:

$$y_{it}=a_{it}\,k_{it}^{\beta 1}h_{it}^{\beta 2} \tag{2}$$

where y_{it} (Y_{it}/ L_{it}) defines the output per worker or average labour productivity which is a function of the amount of capital available per worker (k_{it} =K_{it}/ L_{it}) and human capital per worker ($h_{it} = H_{it}/ L_{it}$). By taking the logarithms and first differencing Equation (2) we obtain:

$$\Delta \ln y_{it} = \alpha + \beta_1 \Delta \ln k_{it} + \beta_2 \Delta \ln h_{it} + \varepsilon_{it} \qquad (3)$$

where $\Delta \ln y_{it}$ is the growth rate in GDP per employed person from period t-1 to period t, while ε_{it} is a stochastic error term capturing all omitted factors. Parameter β_1 accounts for growth in physical capital while β_2 accounts for the growth effects in human capital. According to the Neoclassical economic theory the role of human capital is essential, as once the economy reaches its steady state, physical capital deepening cannot be the sustained source of growth in the economy (Mankiw et al. 1992). This model has been extensively used in the literature but often leading to insignificant effects of human capital (e.g. Islam, 1995, Mason et al., 2012).

It has been debated in the literature whether it is the levels of human capital or the year on year changes in years of schooling is a more important factor affecting growth. For example, based on the works of Romer (1960) and Nelson and Phelps (1966), Benhabib and Spiegel (1994) develop a model where the stock of human capital enters the specification. They assume that human capital influences growth via two channels: by facilitating the catching-up process with the technology leader and by affecting the country's ability to produce new technologies:

$$\Delta \ln y_{it} = \alpha + \beta_1 \ln h_{it-1} + \beta_2 \Delta \ln k_{it} + \beta_3 \Delta \ln y_{it-1} + \varepsilon_{it} \qquad (4)$$

Both, Benhabib and Spiegel (1994) and later Barro and Sala-i-Martin (2007) have found that educational levels are important drivers for growth, but not changes in education level. Other studies such as De la Fuente and Doménech (2006) and Cohen and Soto (2007) find statistically significant evidence on the positive effect of changes in education and economic growth. There are also studies that find strong support for both human capital accumulation and human capital level (Ciccone and Papaioannou 2009).

The latter mechanism is discussed and estimated in Sunde and Visher (2015). These authors identify two specific roles for human capital: on the one hand changes in human capital might accelerate growth by augmenting the existing factors of production, while on the other hand (initial) level of human capital is responsible for the diffusion and adaption of new technologies or through innovation. Estimating the changes but leaving out the initial level of human capital can lead to biased estimates if, as predicted by theory, growth is indeed affected by the initial level of human capital (Sunde and Vischer, 2015). In order to capture both the effect of changes and levels of human capital they extend the Neoclassical model, which can be formally can be written as

$$\Delta \ln y_{it} = \alpha + \beta_1 \Delta \ln h_{it} + \beta_2 \ln h_{it-1} + \beta_3 \Delta \ln k_{it} + \beta_4 \Delta \ln y_{it-1} + \varepsilon_{it} \quad (5)$$

where the human capital parameters are reflected in β_1 and β_2, where $\Delta \ln h_{it}$ refers to the changes in our measures between time t-1 and t, while $\ln h_{it-1}$ is the initial level of human capital in period t-1. The β_1 reflects the effects of human capital in terms of increase in factors (Solow 1956, Lucas, 1988) while β_2 accounts for the growth effects in human capital due to higher adaptiveness to changing environment (Nelson and Phelps 1966). Finally, vector $\ln y_{it-1}$ controls for the initial level of GDP per capita[4].

[4] Other control variables have been included in related models in the empirical literature including growth in physical capital.

3.2 A more general model specification

Although the model we presented earlier is theoretically sound, it does not necessary account for the dynamic relationship between human capital and economic growth. In the second part of the study we further extend our analysis through a model that can account for both short and long run impact of human capital. Several previous studies mention the possibility that the classical first differences growth model is not necessarily a suitable to capture the relationship between human capital and growth, particularly when estimated for a large number of highly heterogeneous countries (Islam 1995, Mason et al. 2012, Sunde and Vischer 2015). This issue has been addressed in Lenkei et al. (2018) with reference to a group of Asian counties. Their study shows that accounting for countries' heterogeneity and cross-sectional dependence in a dynamic panel environment is possibly a more suitable approach for estimating the short and the long-run effects of human capital accumulation. The objective of this section is to extend the methodology of Lenkei et al. (2018) to a wider number of countries including both OECD and emerging markets.

Following the works of Pesaran and Smith (1995), Pesaran et al. (1999) and Eberhardt and Teal (2013); Lenkei et al. (2018) extended the first differences model and proposed that an alternative way of looking at the relationship between human capital and growth is to assume that such relationship follows an autoregressive distributed lag model, which we assume for reasons of simplicity to be of the first order ARDL (1,1,1):

$$\ln y_{it} = \mu_{it} + \delta_{10i} \ln k_{it} + \delta_{11i} \ln k_{it-1} + \gamma_{10\mathrm{i}} \ln h_{it} + \gamma_{11\mathrm{i}} \ln h_{it-1} + \lambda \ln y_{it-1} + \varepsilon_{it} \qquad (6)$$

Equation (5) explains movements in output per worker using the lagged level values of the dependent variable and lagged and current values of all inputs. The previous equation can be also specified as an error correction model (ECM)[5]:

$$\Delta \ln y_{it} = \phi_i (\ln y_{it-1} - \theta_{0i} - \theta_{1i} \ln k_{it-1} - \theta_{2i} \ln h_{it-1}) + \delta_{10i} \Delta \ln k_{it} + \delta_{20i} \Delta \ln h_{it} + \varepsilon_{it} \qquad (7)$$

This specification allows us to analyse both the long run and the short-run impacts of human capital on growth. Specifically, coefficient θ_{2i} captures the stock effect of human capital, or its long-run impact, while the effect of short run adjustments is controlled for by the first difference coefficient δ_{20i}.

Fixed capital per worker is also allowed to affect growth both in the short and in the long run, via the coefficients θ_{1i} and δ_{10i}, respectively. The coefficient ϕ_i is the speed of adjustment to the long-run equilibrium. Equation (7) provides a more general approach for the empirical analysis of the relationship between growth and human capital and allows us to differentiate between various theoretical approaches. In addition, given the availability of a panel of countries with a long-time dimension, this modelling framework will take full advantage of the characteristics of the datasets.

Education versus education inequality

Thomas et al. (1999) argue that the distribution of education (and human capital) is also an important factor for growth models because education is only partially tradable. If an asset is not completely tradable, then the marginal product of the asset across individuals is not equalised, and there is an aggregation problem. In this case,

[5] See Appendix 1 for steps to work out equation (7).

aggregate production functions not only depend on the average level of the asset but also on its distribution.

To measure the impact of education inequality on growth in a dynamic model we apply a similar approach as earlier. Output per worker is estimated using the lagged level values of the dependent variable and lagged and current values of physical capital and the educational GINI coefficient:

$$\ln y_{it} = \mu_{it} + \delta_{10i} \ln k_{it} + \delta_{11i} \ln k_{it-1} + \gamma_{10i} \ln \mathrm{GINI}_{it} + \gamma_{11i} \ln \mathrm{GINI}_{it-1} + \lambda \ln y_{it-1} + \varepsilon_{it} \quad (8)$$

Alternatively, in an error correction model specification Equation (8) can be re-stated as:

$$\Delta \ln y_{it} = \phi_i (\ln y_{it-1} - \theta_{0i} - \theta_{1i} \ln k_{it-1} - \theta_{2i} \ln \mathrm{GINI}_{it-1}) + \delta_{10i} \Delta \ln k_{it} + \delta_{20i} \Delta \ln \mathrm{GINI}_{it} + \varepsilon_{it} \quad (9)$$

Overall, we would expect to see that education inequality has a strong negative statistical relationship with per capita GDP growth which is in line with previous studies (Thomas et al. 1999; Lopez et al 1999).

4. Descriptive analysis

There are two main sources of macroeconomic databases that we utilise in this study. The first one is the Penn World Table (PWT) version 9.0 (Feenstra et al. 2015), that provides data for all countries in the sample on real GDP at current PPPs and capital stock at current PPPs (both in mil. of 2005 USD). Secondly, we compute GDP per worker by supplementing the PWT data with employment series from The Conference Board (2018). We use GDP per worker as it is conceptually more appropriate in growth-accounting regressions and reflects better countries' productive capacity (Pritchett 2001).

We contrast three different measures of human capital. The first measure is the Barro and Lee (2013) (BL) human capital variable which is the most commonly used measure in the economic growth literature (e.g. Islam 1995, Krueger and Lindahl 2001). It provides data on the average years of education for the population aged 15 and above. Observations are available every 5 years for 146 countries between 1950 and 2010. The gaps between observations were filled using linear interpolation. Since the data is only available until 2010, in the analysis we will extrapolate this variable until 2015.

The second human capital indicator is the Cohen and Soto (2007) series (CS) which has been updated by Cohen and Leker (2014). Just like the BL dataset it displays the average years of education for the population aged 15 and above. However, compared to BL it only provides educational attainment data every 10 years over the period of 1960 and 2020 for 95 countries. The key differences between the BL and CD datasets are that the CS use more census observations, employ a different approach to extrapolate missing data, and change values they consider implausible (Ciccone and Papaioannou 2009).

Lastly, in addition to the BL and CS human capital indicators, the series from the Penn World Table (PWT) is also used. The PWT 9.0 human capital index series is available for 150 countries over 1950 to 2014 period and combines data from both BL and CS correcting some of their shortcomings.[6]

[6] It is a challenge to construct data on the average years of schooling in the population. It requires combining information from population censuses with information on school enrolment with significant inconsistencies in classification systems across counties. The PWT presents a more systematic comparison of data from these censuses.

As discussed earlier, the literature argues that instead of the average years of schooling, data on education quality better captures the impact of human capital on growth as it provides a more accurate representation of the education systems across countries (Hanushek and Woessman 2011)[7]. In order to analyse the effect of human capital quality, we make use of the PISA student scores of cognitive abilities. The Programme for International Student Assessment (PISA) is a two-hour test for 15-year-old students conducted every 3 years. It aims to evaluate education systems worldwide by testing students' skills and knowledge in maths, science and reading comprehension. In the 2015 wave data was collected in 70 countries however there are only 55 countries where the test was conducted every occasion since its inception. The PISA tests are considered to be competency based tests. More specifically, it evaluates education systems by assessing how can students apply their obtained knowledge in real life situations at the end of their compulsory education and be equipped for adult life.

In order to measure inequality in educational attainment across countries and time we make use of the educational GINI coefficient (van Leeuwen and van Leeuwen-Li 2015). The Gini of the spread of education in the total population aged 15 years and older is available annually for the period 1850-2010 (the time period varies by country)[8]. This variable was also extrapolated until 2015.

[7] There are also critiques of the international cognitive tests in the literature. Some of the most mentioned critiques include outcome of the tests has no consequence on their school attainment, students have little incentives to provide correct answers and that they are only based on a subsection of the curriculum including science and maths (Pistaferri 2011).

[8] See Appendix 2 for the detailed description on constructing this coefficient.

In the first part of our analysis, in order to make our results comparable across different measures we restrict all our macro and human capital variables to PISA participating studies between 2003 and 2015. Table 1 presents summary statistics of all variables used in our empirical analysis.

Table 1. Summary statistics – Part 1

Variable	Obs.	Mean	SD	Min	Max
Real GDP per worker (Y/L)	649	64.6	34.8	5.2	240.2
Real capital stock per worker (K/L)	649	248.8	136.7	13.2	640.5
Barro and Lee: avg. years of schooling	649	10.4	1.7	6.2	15.0
Penn World Table: avg. years of schooling	649	3.0	0.4	2.0	3.7
Cohen and Soto: avg. years of schooling	451	10.5	2.0	5.2	13.6
PISA mathematics	649	465.9	56.4	284.5	573.5
PISA reading	649	465.7	51.0	272.5	565.4
PISA science	649	472.3	52.6	289.8	569.5
Education Inequality	539	18.4	8.4	4.0	45.9

Note: Sample restricted to PISA participating countries between 2003 and 2015

The shortcoming of the PISA test is that it is being conducted in mainly developed and some emerging countries. However, it is estimated that the return to educational quality maybe even larger in emerging economies (Hanushek and Woessmann 2007), although this effect cannot be suitably captured by this data set. The other issue is that the test is relatively young (maximum of 5 observations for each country until 2015).

Figures 1-6 display the correlation between the changes in human capital and changes in GDP per worker between 2003 and 2015. The three human capital measures of average years of schooling (Figures1-3) suggest a positive but quite weak relationship, while the three PISA test measures (Figures 4-6) indicate a positive and relatively stronger correlation between our variables of interest. It is also worth noting that the data tend to be more dispersed along the trend line when using the average years of schooling measures (Figures 1-3), while the PISA measures (Figures 4-6) are less dispersed. This suggests that on average quantitative factor such as average years of schooling differ more across education systems than the qualitative aspects. These figures are in line with other studies that find similar correlation between changes in human capital and economic growth (Hanushek and Woessman 2007).

All six figures also show the presence of outliers. For example, on Figures 1 & 2 we observe that Kazakhstan has experienced the largest growth and one of the smallest changes of average years of schooling, while on Figures 4-6 the country's above average growth in output per worker was coupled with significant improvements in the country's PISA scores. This suggest that the Kazakh policymakers over the observed period were investing in the enhancement of schooling quality and development of students' cognitive skill, rather than expansion of schooling attainment. We observe somewhat similar patterns for Malaysia and Argentina i.e. modest improvement in average years of schooling but substantial gains in the quality of schooling. However, for example Qatar has produced large changes in both areas indicating that that the government has been focusing on allocating resources to education improving both accessibility to basic education in all parts of the society while at the same time improving the quality of their system.

Figures 1-6. Relationship between GDP per worker and human capital

Fig.1

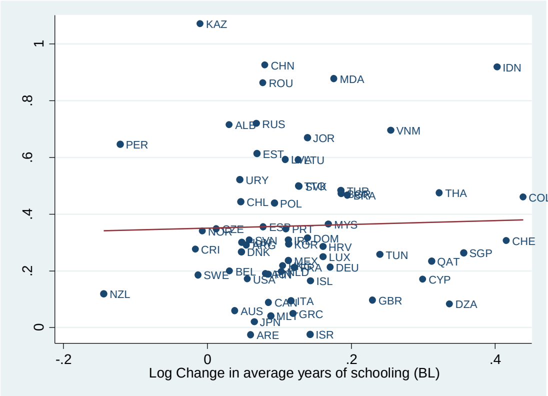


Fig.2

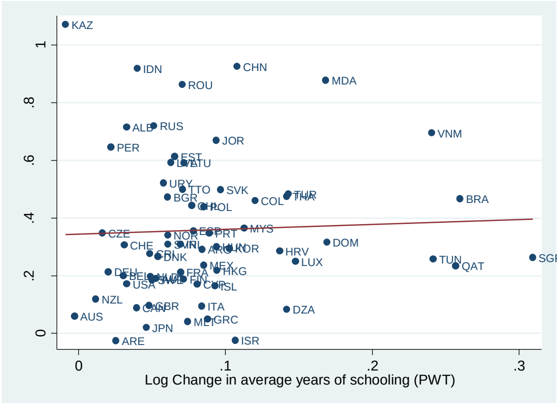


Fig.3

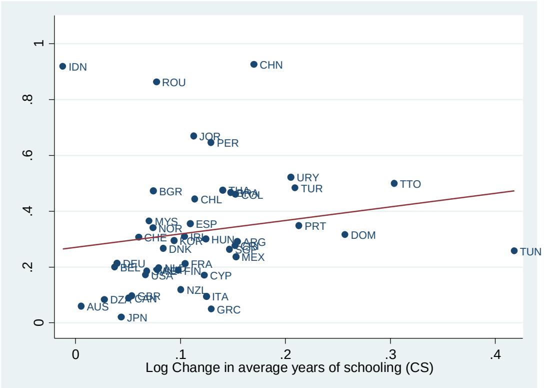


Fig.4

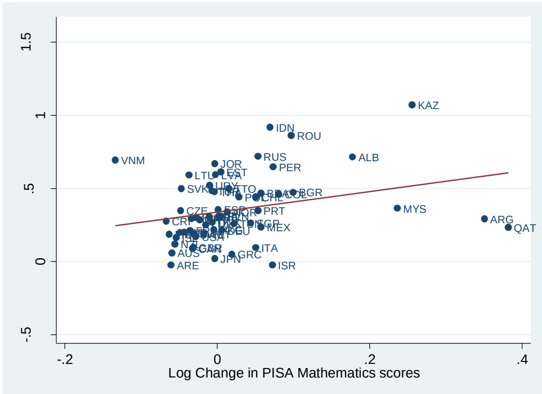


Fig.5

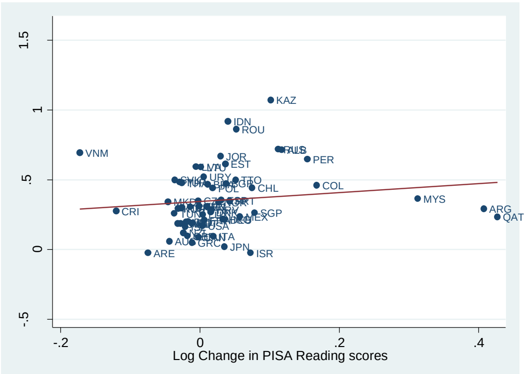
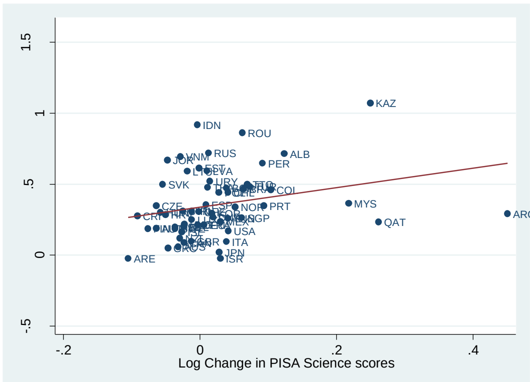


Fig.6



Following the works of Benhabib and Spiegel (1994) and Sunde and Vischer (2015), in Appendix 3 we test whether there might be a potential bias in our estimates if we only account for just one channel in which human capital affects growth. On Figures (A1-A6) we plot the relationship between initial levels of human capital in 2003 and subsequent changes in these variables between 2003 and 2015 using average schooling measures (BL, PWT, CS) and PISA scores (maths, science reading). All figures display a strong and negative correlation between the variables. This suggests that the higher the initial level of human capital the lower the subsequent percentage change in average years of schooling. These results are in line with previous findings, suggesting that due to the negative correlation between the initial level and changes in human capital, if any of them are omitted from the regression the estimates should be biased downwards.

We again observe that the average years of schooling measures of human capital are more dispersed around the trend line, while there is a clearer relationship between the variables using PISA. Interestingly Singapore, that topped the PISA tables in 2015, has experienced some of the largest changes in average years of schooling (Figure A1 & A2), but they have not made substantial gains in PISA scores over the observed time period (Figures A4-A6). This is because the country already had above average initial achievement scores on the PISA test and most probably the government was rather focusing resources on improving average attainment for the wider society.

On Figures A1-A2 we obtain some outlier observations as countries such as Argentina, Kazakhstan and Malaysia and Qatar have had the smallest initial level of PISA scores, however they have seen the largest improvement. However, besides improving on their education quality, Qatar has also seen large gains in their average years of schooling figures from the low initial levels.

Finally, Figures 7-8 present the correlation between changes in educational inequality and changes in GDP per worker and initial inequality, respectively. Both plots display a positive correlation between our variables, although neither of the correlations are robust. Figure 7 suggests that on average changes in educational inequality are positively correlated with changes in GDP per worker, while Figure 8 indicates that higher initial inequality is positively correlated with higher percentage change in subsequent educational inequality, i.e. inequality appears to be persistent. This finding is slightly puzzling because it does not meet our prior expectations and it contradicts parts of the literature (e.g. Castello and Domenech 2002).

Figures 7-8. Relationship between inequality and
GDP in per worker and initial inequality

Fig.7.

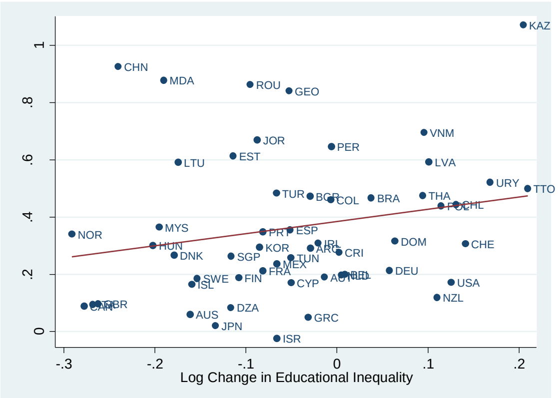
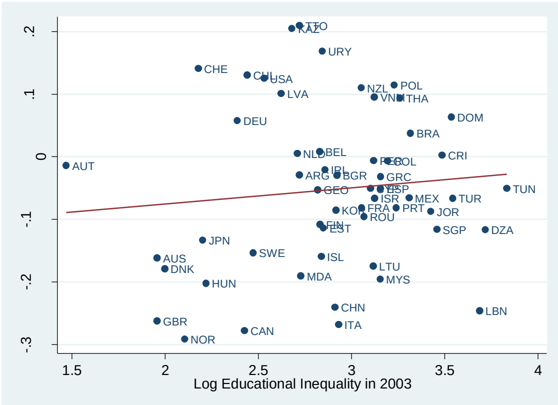


Fig.8.



One may argue that since the human capital inequality was linearly interpolated since 2010 we might obtain biased correlation estimates. In Appendix Figure A7 we re-created Figure 7 using the average educational inequality between 2003 and 2015 instead of changes in inequality. However, we obtain results similar to those shown in Figure7, suggesting that countries with higher average educational inequalities have seen the largest changes in output per worker between 2003 and 2015. An interesting example is China that has one of the highest educational inequality in the world and also has seen the smallest changes in these figures but produced one of the largest changes in output per worker during the period under analysis.

5. Results and Discussion

5.1 Traditional Model

Table 2 presents the estimates of our average schooling measures (BL, PWT, CS) in growth regressions. All estimates on changes in human capital are negative and insignificant when we do not account for initial levels of human capital (Columns 1,4 and 7). These findings are in line with previous findings reported in the literature (e.g. Benhabib and Spiegel 1994; Sunde and Visher 2015). The specification with the initial average years of schooling (Columns 2,5,8) provides positive and significant effect of initial human capital on growth which is consistent with previous studies (e.g. Makiw et al 1992). This suggests that a higher initial level of average schooling provides a larger change in GDP per employed person. Once we account for both channels (Columns 3,6,9), the $\Delta \ln h$ estimates turn positive for BL and CS measure but not for the PWT, however all three coefficients remain insignificant. This finding is in line with the conjecture that weak empirical evidence obtained in previous studies might have been a result of omitted variable bias. Overall, our results suggest that the levels are more important factors in affecting productivity as compared to changes in human capital.

We replicate the previous table in Table A1 (Appendix) using short instead of long differences in the regression specifications between 2003 and 2015. We obtain similar results as in Table 2, however the coefficients on the initial level of human capital (ln h_{t-1}) become insignificant.

Table 2. Growth Regressions using average schooling HC measures

	Barro and Lee			Penn World Table			Cohen and Soto		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Δ lnh	-0.134		0.159	-0.461		-0.050	-0.225		0.131
	(0.190)		(0.169)	(0.409)		(0.429)	(0.320)		(0.419)
Ln h_{t-1}		0.270**	0.328**		0.306**	0.292*		0.176	0.211
		(0.124)	(0.126)		(0.144)	(0.148)		(0.119)	(0.153)
Δ lnk	0.095	0.099	0.108	0.100	0.116	0.115	0.082	0.097	0.105
	(0.091)	(0.092)	(0.097)	(0.089)	(0.093)	(0.096)	(0.108)	(0.111)	(0.110)
Ln y_{t-1}	-0.262***	-0.293***	-0.296***	-0.266***	-0.288***	-0.288***	-0.262***	-0.296***	-0.299***
	(0.035)	(0.042)	(0.041)	(0.037)	(0.041)	(0.040)	(0.037)	(0.046)	(0.047)
Constant	1.330***	0.824***	0.682***	1.369***	1.088***	1.105***	1.321***	1.026***	0.940***
	(0.164)	(0.218)	(0.252)	(0.179)	(0.143)	(0.220)	(0.175)	(0.217)	(0.346)
Observations	65	65	65	65	65	65	45	45	45
R-squared	0.666	0.699	0.703	0.676	0.693	0.694	0.677	0.693	0.694

Notes: Dependent variable: annualized long difference in log GDP/worker (2003–15). Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Moreover, in Table 2 using long differences we find that the change in physical capital is never statistically significant, whereas in Tables A1 (Appendix) using short differences the physical capital variables turn significant. This suggests that there are limitations with these specifications, which will be addressed in the next section where we use dynamic panel techniques.

In order to investigate the consequences of applying measures of human capital quality rather than quantity the same analysis is performed using student achievement scores on the PISA test as proxies for human capital. We differentiate between achievement scores in mathematics, reading and science tests. As mentioned earlier, the PISA test was collected tri-annually since 2003. There are two ways of applying the traditional model on our data. Firstly, we can take long differences over the periods between 2003 and 2015. Secondly, we can use linear interpolation to fill up the gaps between PISA observations and use the short differences method. In order to make different human capital measures comparable we use the same time periods and the same group of PISA participating countries for the analysis[9]. The only drawback of choosing this time period is the difficulty in comparing our estimates with earlier results in the literature.

Results in Table 3 show that when accounting for both channels of changes and initial level of human capital quality a similar picture emerges as earlier. The changes in PISA scores across all three disciplines are positive, however it only gets significant for mathematics scores (Column 3). Despite the sample of countries used are slightly different, the impact of the PISA human capital measure is substantially larger than either the BL, PWT or CS measures, however the effect is slightly weaker (statistically significant only at the 10% level). These results confirm to some extent previously reported findings in the literature regarding the importance of education quality for growth. For example, Hanushek and Woessmann (2012), using average standardised maths and science scores in international tests as a proxy for cognitive skills, find that differences in cognitive skills lead to economically significant differences in economic growth.

[9] The countries included are Albania, Argentina, Australia, Austria, Belgium, Brazil, Bulgaria, Canada, Chile, Macao, Colombia, Costa Rica, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hong Kong, Hungary, Iceland, Indonesia, Ireland, Israel, Italy, Japan, Jordan, Kazakhstan, Latvia, Lithuania, Luxembourg, Malaysia, Mexico, Montenegro, Netherlands, New Zealand, Norway, Peru, Poland, Portugal, Qatar, Romania, Russian Federation, Singapore, Slovak Republic, Slovenia, South Korea, Spain, Sweden, Switzerland, Thailand, Trinidad & Tobago, Tunisia, Turkey, United Arab Emirates, United Kingdom, United States, Uruguay and Vietnam.

In Table A2 we perform the same regressions using short difference regressions however the coefficient on the changes in PISA mathematics scores is not weakly significant anymore. This suggests that the method of analysis has a large impact on findings.

Table 3. Growth Regressions using PISA test scores HC measures

	PISA Mathematics			PISA Reading			PISA Science		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Δ lnh	0.398		0.937*	0.293		0.533	0.296		0.448
	(0.287)		(0.542)	(0.222)		(0.394)	(0.333)		(0.560)
Ln h_{t-1}		-0.085	0.409		-0.108	0.201		-0.123	0.131
		(0.146)	(0.258)		(0.130)	(0.224)		(0.174)	(0.309)
Δ lnk	0.027	0.058	-0.030	0.048	0.059	0.033	0.048	0.065	0.034
	(0.119)	(0.114)	(0.131)	(0.112)	(0.112)	(0.116)	(0.117)	(0.110)	(0.121)
Ln y_{t-1}	-0.254***	-0.249***	-0.293***	-0.256***	-0.246***	-0.274***	-0.251***	-0.245***	-0.262***
	(0.032)	(0.038)	(0.046)	(0.035)	(0.036)	(0.043)	(0.034)	(0.037)	(0.047)
Constant	1.301***	1.806**	-1.053	1.301***	1.928**	0.136	1.289***	2.024*	0.524
	(0.142)	(0.873)	(1.454)	(0.154)	(0.793)	(1.296)	(0.150)	(1.048)	(1.787)
Observations	59	59	59	60	60	60	59	59	59
R-squared	0.653	0.632	0.673	0.638	0.626	0.642	0.642	0.633	0.643

Notes: Dependent variable: annualized long difference in log GDP/employed person (2003–15).
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Finally, in Table 4 we complement our discussion on human capital with regression estimates on the impact of educational inequality (as measured by educational GINI) on growth. The variable on changes in inequality is positive but insignificant, broadly the same magnitude as the BL and CS measures. This result is consistent with Foldvari and van Leeuwen (2011) who also find that the GINI coefficient of education yields an insignificant coefficient. However, as opposed to the previous two tables the variable on the initial level of inequality is negative and weakly significant suggesting that higher levels of inequality are related to smaller subsequent changes in GDP per employed person. In Table A3 we replicated the human capital inequality regressions using short difference estimates. Despite all our coefficients have the same sign, we do not overserve a significant effect of initial inequality in 2003. This suggests again that the results are sensitive to the specification used to analyse the impact of human capital.

Table 4. Growth Regressions using Education Inequality measures

	Educational Inequality		
	(1)	(2)	(3)
Δ lnGini	0.136		0.166
	(0.184)		(0.179)
Ln $Gini_{t-1}$		-0.078*	-0.082*
		(0.044)	(0.045)
Δ lnk	0.097	0.078	0.108
	(0.101)	(0.095)	(0.101)
Ln y_{t-1}	-0.255***	-0.282***	-0.272***
	(0.031)	(0.037)	(0.033)
Constant	1.283***	1.602***	1.579***
	(0.133)	(0.232)	(0.218)
Observations	54	54	54
R-squared	0.658	0.673	0.679

Notes: Dependent variable: annualized long difference in log GDP/worker (2003–15). Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

5.2 Dynamic specification

Although the analysis so far is theoretically sound it does not fully capture the dynamic effect of human capital on growth. This may also explain the lack of robust results. Therefore, we extend our econometric approach to complement our discussion on the different channels that may impact growth.

The lack of observations in both cross sectional and time series dimensions make the analysis challenging in the second part of the study where we look at cross sectional dependence and heterogeneity in a dynamic panel setting. As such, in this section we only focus on the average schooling (BL, PWT, CS) and educational inequality measures, which allows us to broaden the number of participating countries and to extend the time period under analysis between 1960 and 2015 in our growth regressions.

Besides running the regressions using the full sample, due to the longer time dimension and a greater number of countries available, the data also allows us to create groups of countries at similar stages of economic development (e.g. OECD and Emerging Markets (EMs) from Asia, South America and Africa)[10]. This can help to alleviate the relevance of the heterogeneity problem. However, assuming the same returns to factor inputs even within these groups of countries is a strong assumption. The use of standard panel data techniques can only partially address this issue under the assumption that cross-country differences are fixed over time. A much more flexible estimation strategy is to allow the technology parameters to vary by estimating separate equations for each country and then derive the mean of individual estimates, as in the Mean Group (MG) estimator (Pesaran and Smith,1995; Pesaran et al. 1999; Bond and Eberhardt 2013).

In addition, another issue is related to the assumption of cross-sectional independence. This means that in the empirical analysis each country (cross-sectional unit) is treated as a separate entity with no impact on other countries' technologies. So far this issue has been mainly neglected in the human capital literature, with a notable exception of Lenkei et al. (2018). However, this is a strong assumption given increasing interconnections across economies, common shock such as financial crises, globalisation and international spill-overs, including human capital spill-over and productivity spill-overs (Engelbrecht 1997). This is especially true when using panel dataset with long-time dimensions. This increases a possibility of bias in our coefficient estimates. Even if spill-overs are not present between

[10] This distinction would support the view that human capital accumulation has differing impact on different nations depending on their level of development (Vandenbussche et al 2006).

countries, other unobserved common effects can be accounted for by controlling for cross-sectional dependence (Eberhardt et al., 2013).

For the estimation of the dynamic model (Equation 6), we adopt the Pesaran et al. (1999) Pooled Mean Group estimator method. This allows short run coefficients and error variances to vary across countries, while imposing common long-run coefficients. This means that technology parameters will converge to common values in the long-run, while they differ in the short run, a reasonable assumption when countries have access to similar technologies. Imposing common long run coefficients increases the efficiency of the estimates and it is therefore preferred in this setting. This estimation framework also allows us to test directly for the presence of co-integration, based on the significance of the error correction term (the speed of adjustment). To account for cross sectional dependence, we extend the PMG estimator to include the common factors in the specification of the short-run dynamics (Binder and Offermanns, 2014).

In our study the adoption of the ARDL modelling framework will produce consistent estimates even in the presence of endogenous regressors, as discussed in Pesaran and Shin (1999). Another issue is related to cross sectional independence i.e. each cross-section unit is treated as an isolated entity without impact on other countries' technologies. Omitting factors that capture the effects of common shocks can make our estimates biased and inconsistent. By adding controls for cross-sectional dependence, we account for an additional source of omitted variable bias.

In Table 5 we present the results of the estimation for the first differences model under the assumption of cross-sectional independence and exogeneous regressors between 1960 and 2015. The results provide mixed evidence. The three human capital measures are negative but while the Cohen Soto measure is statistically significant, the Barro-Lee and PWT measures remain insignificant. We replicate the analysis in Table A4 restricting the samples to OECD and EMs, however we do not find any significant effect for any of the three human capital indicators in the two sub-samples. The general lack of significant results is in line with our previous findings (Table 2) and other findings in the human capital literature (e.g. Benhabib and Spiegel 1994; Sunde and Vischer, 2015). Furthermore, negative results were also found in Pritchett (2001) using an 'educational capital' measure (based on Barro Lee data) between 1960 and 1987. He justifies the presence of this negative relationship with slow growth in the demand for educated labour and failed education systems that may provide few skills.

Table 5. First Differences models: common technologies and cross-sectional independence (1960-2015)

	(1) Barro Lee	(2) Cohen Soto	(3) PWT
Δlnk_{it}	0.487***	0.478***	0.485***
	(0.055)	(0.055)	(0.053)
Δlnh_{it} (Barro-Lee)	-0.053		
	(0.092)		
Δlnh_{it} (Cohen-Soto)		-0.012***	
		(0.004)	
Δlnh_{it} (PWT)			-0.459
			(0.348)
Constant	0.043***	0.042***	0.042***
	(0.015)	(0.016)	(0.016)
Observations	4,601	3,561	4,825
R-squared	0.111	0.108	0.108

Note: Dependent variable: value added per worker in log first differences. Estimation is based on a fixed effect estimator and all specifications include time and country dummies. Standard errors in brackets. * = significant at 10%, ** = significant at 5%, *** = significant at 1%.

There are two important concerns with the results presented in Table 5. Firstly, the assumption of cross-sectional independence and secondly, the assumption that there are the same technology coefficients across all the countries. In Table A5 we re-estimate the model by controlling for cross-sectional dependence using mean group estimator (Pesaran 2006). Similar to the models with cross sectional independence we obtain mixed results. However, the estimates in Table A5 are overall preferred compared to Table 5, because by controlling for CSD, we are possibly accounting for factors that can affect the impact of human capital, such as corruption, militarism, religious and political conflicts and other institutional factors.

Next, we present our analysis based on a fully dynamic model, which provides an alternative way of modelling human capital stock and human capital accumulation. These models are more complex as they use the ECM framework, as discussed earlier, it allows us to capture the short and long run-impact of human capital on growth (Pesaran 2009). The estimator imposes the same long-run coefficient on the countries included in the regression but makes the short-run effects to be specific to each country. Results presented in Tables 6 are based on the PMG estimator (Pesaran et al. 1999), where Columns 1-3 present results based on the assumption of cross-sectional independence, while in Columns 4-6 the correction for CSD is included in the short-run specification (Binder and Offermanns 2014).

In all model specifications the error correction term is significant and has a negative sign as expected. This confirms the validity of the framework and the presence of a long-run equilibrium. On average, the EC term predicts a similar rate of adjustment across all models. It can be interpreted as each year our sample of countries will approximately close 10% of the gap between short-run and long run equilibrium. The growth in output per worker is positively and significantly related to the long-run coefficients on human and physical capital. This is true for the BL, CS and PWT measures. The only insignificant results are the BL coefficients in the no CDS specification (Column 1).

As both the long-run coefficients on physical capital per worker and the error correction term are statistically significant, it implies that the models presented in the previous sections suffer from the omitted variable bias. The changes in output per capita are positively related to changes in physical and human capital, suggesting that human and physical capital accumulation played an important role in supporting growth in our sample of countries in the long-run (short run coefficient on education are either negative or insignificant – or both)[11].

In Tables 4-6 the CS and PWT coefficients are larger compared to the BL measure of human capital. This suggests that controlling for the differential returns to education can considerably increase the role that human capital plays in growth models. This is consistent with other studies in the human capital literature (e.g. Mason et al., 2012).

We replicate the results in Appendix Table A6 and A7 using OECD and Emerging Market samples respectively. The findings are unchanged, suggesting that human capital investments have a broadly positive and significant effect on productivity in the long-run, but not in the short-run. There are some slight differences however between the two groups of countries. For instance, both the BL and CS estimates are smaller for OECD compared to emerging economies in the long-run, which suggest

[11] A reason explaining the negative short-run impact is that initial growth inhuman capital can be costly for developing countries, however in the long run the rates of accumulation is expected to outweigh these effects.

that increasing average levels of education should be crucial in promoting growth in emerging economies.

If we compare the results with related studies in the literature, we find that improving the model specification by accounting for cross-sectional dependence and heterogeneity we obtain stronger effects of human capital on economic growth in both in developed and developing economies as compared to previous studies (e.g. Mankiw et al. 1992; Collins and Bosworth 1996; Pritchett 2001; Baldacci et al. 2008; Park 2012). For instance, the findings of Collins and Bosworth (1996) implies a range of values for the human capital variables between 0.16 and 0.25. Similarly, Baldacci et al. (2008) estimation for the human capital values is approximately 0.1. Our estimates are closer to the results of Sunde and Vischer's (2015) who estimated human capital coefficients between 0.3 and 0.6 for a sample of 90 countries, although the methods used in this study differ.

Although direct comparison of the results of these studies is challenging given the differences in the composition of samples, it is clear from the results based on the dynamic model specification produces larger effects for human capital compared to estimates in previous studies. This implies that some studies in the past may have underestimated the importance education in promoting economic growth, whereas our study provides a stronger argument towards policies targeting increased investment in education. There are other factors that are also important in promoting economic growth including investments in research and development (Griffith et al. 2004), product quality and variety (Funke and Strulik, 2000), and access to finance (Levine and Zervos, 1998); however, improvements in human capital is a prerequisite for achieving other targets.

Table 6. Pooled mean group estimator of the relationship between human capital and growth. All countries. (1960-2015)

	No correction for CSD ARDL(1,1,1)			Correction for CSD ARDL(1,1,1)			Correction for CSD ARDL(2,2,2)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
lnk_{it}	0.656***	-0.554***	0.298***	0.618***	0.386***	0.476***	0.606***	0.325***	0.640***
	(0.020)	(0.125)	(0.028)	(0.024)	(0.031)	(0.022)	(0.023)	(0.034)	(0.019)
lnh_{it} (Barro-Lee)	-0.025			0.117***			0.177***		
	(0.023)			(0.036)			(0.031)		
lnh_{it} (Cohen-Soto)		1.905***			0.269***			0.837***	
		(0.170)			(0.023)			(0.054)	
lnh_{it} (PWT)			1.394***			1.075***			0.458***
			(0.079)			(0.071)			(0.062)
EC	-0.090***	-0.043***	-0.104***	-0.140***	-0.097***	-0.141***	-0.159***	-0.100***	-0.162***
	(0.010)	(0.007)	(0.009)	(0.011)	(0.012)	(0.011)	(0.012)	(0.012)	(0.013)
Δlnk_{it}	0.404***	0.451***	0.399***	0.416***	0.393***	0.364***	0.535***	0.544***	0.484***
	(0.062)	(0.061)	(0.070)	(0.064)	(0.068)	(0.071)	(0.075)	(0.096)	(0.078)
Δlnh_{it} (Barro-Lee)	-0.102			0.565			0.252		
	(0.546)			(0.430)			(0.427)		
Δlnh_{it} (Cohen-Soto)		-0.552			-0.566*			0.673	
		(0.373)			(0.339)			(0.503)	
Δlnh_{it} (PWT)			-2.631***			-1.492**			-1.314
			(0.896)			(0.714)			(0.911)
Constant	0.039***	0.093***	0.089***	0.043***	0.072***	0.037	0.022	-0.008	-0.010
	(0.008)	(0.016)	(0.011)	(0.014)	(0.015)	(0.028)	(0.014)	(0.012)	(0.034)
Observations	5,283	4,137	5,538	5,283	4,137	5,538	5,236	4,057	5,486

Notes: Dependent variable: value added per worker in log first differences. EC stands for Error Correction coefficient. Specifications (4)–(9) are augmented with cross section averages (not reported). Standard errors in brackets. * = significant at 10%, ** = significant at 5%, *** = significant at 1%.

5.3 Educational Inequality in the Dynamic Model

In this section we address the issue of whether large inequalities in human capital distribution within a nation plays an important role in growth. If the ability of a nation's population is normally distributed, then a skewed distribution of educational opportunities will lead to welfare losses. As such, an equitable distribution of human capital constitutes a pre-requisite for substantial gains in productivity (Thomas et al. 2002).

For this analysis we apply the educational GINI coefficient by van Leeuwen and van Leeuwen-Li (2015), originally proposed by Thomas et al. (2002). Table 7 presents the results of the first differences regressions where columns (1)-(3) assume common technology and cross-sectional independence, while columns (4)-(6) assume common technology and cross-sectional dependence. Both GINI coefficients have the expected negative sign but remain insignificant. We re-run these regressions in Table 8 also controlling for the BL, PWT and CS measures of HC but the educational GINI coefficients are still non-significant with the same sign as without HC controls. In Table A8 using OECD and emerging market sub-samples the GINI coefficients are still non-significant, but their sign turn positive for OECD countries suggesting that inequality impacts EMs more negatively than developed markets.

Table 7. First differences model: Educational inequality as a determinant of growth (1960-2015)

	(1) Cross-sectional Independence	(2) Cross-sectional Dependence
Δlnk_{it}	0.458***	0.374***
	-0.062	-0.062
$\Delta lnGINI_{it}$	-0.009	-0.173
	-0.03	-0.187
Constant	0.040***	-0.009
	-0.015	-0.009
Observations	4,242	4,242
R-squared	0.107	
Number of id	101	101

Note: Dependent variable: value added per worker in log first differences. Estimation is based on a fixed effect estimator and all specifications include time and country dummies. Columns (1)-(3) assume common technologies and cross-sectional independence, while columns (4)-(5) assume common technologies and cross-sectional dependence. Standard errors in brackets * = significant at 10%, ** = significant at 5%, *** = significant at 1%.

Table 8. First differences model: Educational inequality as a determinant of growth, controlling for HC measures (1960-2015)

	(1) Barro Lee	(2) Penn World Table	(3) Cohen Soto
Δlnk_{it}	0.475***	0.473***	0.489***
	(0.056)	(0.054)	(0.059)
$\Delta lnGINI_{it}$	-0.007	-0.008	-0.050
	(0.031)	(0.031)	(0.051)
Δlnh_{it} (Barro-Lee)	-0.016		
	(0.092)		
Δlnh_{it} (PWT)		-0.271	
		(0.448)	
Δlnh_{it} (CS)			-0.013***
			(0.004)
Constant	0.042**	0.042***	0.042**
	(0.016)	(0.015)	(0.016)
Observations	3,967	4,146	3,427
R-squared	0.106	0.102	0.108
Number of id	93	97	77

Note: Dependent variable: value added per worker in log first differences. Assuming common technologies and cross-sectional dependence Estimation is based on a fixed effect estimator and all specifications include time and country dummies. Standard errors in brackets * = significant at 10%, ** = significant at 5%, *** = significant at 1%.

In Table 9 we estimate the impact of educational inequality using the pooled mean group estimator method. The same long-run coefficients are imposed across our samples, but short-run adjustments are allowed to be country specific. The long-run elasticity of the educational inequality GINI coefficients ($lnGINI_{it}$) turn negative, while the short run impacts ($\Delta lnGINI_{it}$) are positive (except from Column 2). Using the coefficient in Column (3) we can interpret the long run coefficients as follows: if the long run average educational Gini were to increase 1% it would decrease changes in the long run GDP per worker by 1.2%. The regressions are replicated in Tables A9 for OECD and emerging market samples. We observe generally larger coefficients for emerging economies suggesting that in these countries improvements in terms of education inequality leads to larger growth opportunities. It is important to note however that the results are sensitive to the dynamic specification.

Table 9. PMGE: Educational inequality as a determinant of growth

	(1) No correction for CSD ARDL(1,1,1)	(2) Correction for CSD ARDL(1,1,1)	(3) Correction for CSD ARDL(2,2,2)
lnk_{it}	0.937***	0.512***	0.754***
	(0.101)	(0.025)	(0.029)
$lnGINI_{it}$	-3.794***	-0.008	-1.278***
	(0.354)	(0.042)	(0.038)
EC	-0.034***	-0.091***	-0.094***
	(0.007)	(0.012)	(0.012)
Δlnk_{it}	0.397***	0.459***	0.477***
	(0.066)	(0.077)	(0.088)
$\Delta lnGINI_{it}$	0.463***	-0.113	0.192
	(0.150)	(0.199)	(0.203)
Constant	0.397***	0.052***	0.372***
	(0.074)	(0.013)	(0.055)
Obs.	4,242	4,242	4,220

Notes: Dependent variable: value added per worker in log first differences. EC stands for Error Correction coefficient. Specifications (4)–(9) are augmented with cross section averages (not reported). Standard errors in brackets. * = significant at 10%, ** = significant at 5%, *** = significant at 1%.

The long run negative sign of the GINI coefficient on educational inequality hindering economic growth meets our prior expectations[12]. They highlight the importance of the

[12] Nevertheless, some past studies obtained a negative relationship between equality in education and income per capita (e.g. Gille, 2015).

United Nations (UN) efforts to combat inequality in the word. The UN is currently calling for the adoption of what is known as the Sustainable Development Goals (SDG), which is the successor to the Millennium Development Goals (United Nations Development Programme, 2005). One of the goals of the SDG is to set forth an intergovernmental framework that provides equitable and inclusive education for all and to reduce economic, political and social equality within nations. The results are consistent with the literature suggesting that it is more important to increase the supply of workers in the primary and secondary levels to improve productivity (Self and Grabowski 2004, Keller 2006), as opposed to other studies highlighting that the impact of completed higher education is a more important driver of economic prosperity compared to general accessibility to lower levels of schooling (Castello-Climent and Mukhopadhyay 2013, Castello-Climent et al. 2017).

6. Conclusion

In the last few decades the role of schooling has become a central part of most nations and some international organisations' (e.g. UN, OECD) development strategies (Hanushek and Woessmann 2007). This study provides a discussion on the impact of the quantity and quality human capital and educational inequality on economic growth, with the main objective of bringing some resolution to the mixed evidence and shed light on some of the unresolved issues that still plague the macroeconomic literature. We use different measures of human capital and model specifications and present a range of econometric modelling techniques to capture the channels through which human capital can impact on growth. Firstly, using long differences estimators we do not find that changes in the average years of schooling between 2003 and 2015 significantly affects growth by using either of the three HC measures (BL, PWT, CS). However, initial levels of HC do have an impact in a way that higher levels of initial average years of schooling in 2003 significantly contributes to higher growth over the observed time period. We complement the discussion in the literature on the effects of human capital quantity versus quality when we re-run the same regressions controlling PISA scores capturing the effects of HC quality. Our results provide some evidence that quality, captured by PISA mathematics score, has a somewhat stronger impact as the magnitude of our quality coefficients are larger, although their effect is weakly significant. Furthermore, we investigate the impact of the inequality of educational attainment using the long differences modelling framework. Access to education should be available for the whole population in most counties. If people's abilities are normally distributed, then a skewed distribution of educational opportunities represent welfare losses. Although we observe negative coefficients, they don’t turn significant at any of our modelling frameworks. This is an important issue that requires further investigation.

Given some of the limitations of the long differences model the second part of the study makes use of new developments in panel data techniques. By controlling for the possible presence of cross-sectional heterogeneity and dependence between countries, our results suggest that in the last few decades, changes in human capital stock have played an important role in increasing productivity across countries. While the first differences model does not support the view that investments in education plays a significant role in raising growth in output, the ECM framework, our preferred model, predicts a statistically significant increase in output given a 1% increase in human capital. This is consistent across all three measures of human capital (BL, CS and PWT). In addition, we observe a larger magnitude for human capital coefficients in our sample of developing countries. This is consistent with previous findings in the HC literature (e.g. Lenkei et al 2018). We also find that in the PMG framework the CS and PWT measures human capital measures are larger compared to the BL measure. This is because the CS and PWT adjust information on educational attainment with different returns for primary, secondary and tertiary education, therefore producing larger coefficients. This suggests that accounting for the quality of education gives an even stronger human capital effect. In addition, the dynamic panel data model finds significant evidence that in the long run a more equal distribution of education within societies (measured by educational GINI) leads to higher changes in income per worker.

The findings from this analysis raise important questions for policymakers across the world. For instance, should countries focus resources on increasing the percentage of population accessing basic education or instead improving the educational circumstances for current students. It is a common objective for governments to expand access to education, especially at the basic level. But for any level of efficiency, in order to maintain quality, increased number of enrolments require increased resources. This is especially true for developing countries where increases in enrolments can be accompanied by worsening schooling conditions, lower public school spending and increasing student teacher ratios.

Overall, the study provides evidence that, over the last few decades, human capital has had an impact on economic growth. This is particularly true when using dynamic specification and controlling for cross-sectional heterogeneity and unknown common factors. Our results support the view that governmental policies should focus on allocating national resources to education improving both accessibility to basic education in all parts of the society while at the same time improving the quality of education systems.

In addition, further analysis is required to investigate if our results hold for different sub-sub-samples of countries (e.g. by regions). This is necessary as different environmental factors and the quality of education systems can differ across countries, implying differential effects of human capital. Moreover, the overall economic institutions also have an impact on growth. For instance, the security of the nation, the openness of the economy, a well-defined system of property rights can be considered almost as prerequisites to economic growth. Without these efficiencies, skills and education may not have the desired impact on economic development and we cannot presume that any spending is a productive investment. These factors all have implications for policy interventions that should consider countries' specific needs.

Another area worth exploring in further studies is the differential impact of primary, secondary and tertiary education on growth (Lenkei et al 2018). This would be able to further address the issue whether policies should be directed towards increasing investments in primary and secondary education or focusing on a minority educated at the tertiary level. A further extension to macro growth models is the potential incorporation of non-cognitive. Non-cognitive skills are a broad category of metrics encompassing personality, behaviours and socio-emotional skills (Lundberg 2015). However, one of the shortcoming of these measures is that they are difficult to find for a large number of countries.

References

Barro, R.J., Sala-i-Martin, X. (1997) Technological diffusion, convergence, and growth. Journal of Economic Growth, 2: 1–26.

Barro, R.J., Lee, J.W. (2013) A New Data Set of Educational Attainment in the World, 1950-2010. Journal of Development Economics, 104: 184-198.

Basu, A.M. (2002) Why does education lead to lower fertility? A critical review of some of the possibilities, World Development, 30(10): 1779–1790.

Bedard, K., Ferrall, C. (2003) Wage and Test Score Dispersion: Some International Evidence. Economics of Education Review, 22: 31-43.

Behrman, J., Ross, D., Sabot, R. (2008) Improving Quality versus Increasing the Quantity of Schooling: Estimates of Rates of Return from Rural Pakistan, Journal of Development Economics, 85 (1-2): 94–104.

Benhabib, J., Spiegel, M.M. (1994) The role of human capital in economic development evidence from aggregate cross-country data. Journal of Monetary Economics, 34: 143–173.

Binder, M., Offermanns, C.J. (2014) International Investment Positions and Exchange Rate Dynamics, Discussion Paper N. 2014/23. School of Business & Economics, Free University Berlin.

Birdsall, N. (1996) Public Spending on Higher Education in Developing Countries: Too Much or Too Little? Economics of Education Review, 15(4): 407-419.

Birdsall, N., Londoño, J.L. (1997) Asset Inequality Matters: An Assessment of the World Banks Approach to Poverty Reduction. American Economic Review, 87 (2): 32-37.

Blau, F., Kahn, L. (2005) Do Cognitive Test Scores Explain Higher US Wage Inequality? Review of Economics and Statistics, 87: 184-193.

Bond, S., Eberhardt, M. (2013) Accounting for Unobserved Heterogeneity in Panel Time Series Models. Nuffieldeld College, University of Oxford, mimeo.

Borgonovi, F., Burns, T. (2015) The Educational Roots of Trust, OECD Education Working Papers, No. 119, Paris, OECD Publishing.

Cantoni, D., Yuchtman, N. (2014) Medieval universities, legal institutions, and the commercial revolution. Quarterly Journal of Economics, 129(2): 823–887.

Castelló-Climent, A., Doménech, R. (2002). Human capital inequality and economic growth: some new

evidence. The Economic Journal, 112(478): 187-200.

Castelló-Climent, A., Doménech, R. (2014) Human Capital and Income Inequality: Some Facts and Some Puzzles. BBVA Working Paper No. 12/ 28.

Castelló-Climent, A., Chaudhary, L., Mukhopadhyay, A. (2017) Higher Education and Prosperity: From Catholic Missionaries to Luminosity in India. The Economic Journal, 128 (616): 3039-3075.

Castelló-Climent, A., Mukhopadhyay, A. (2013) Mass education or a minority well educated elite in the process of growth: the case of India. Journal of Development Economics, 105: 303–320.

Checci, D. (2001) Education Inequality and Income Inequality. STICERD Distributional Analysis Research Programme Discussion Paper No. 52.

Chudik, A., Pesaran, M.H., Tosetti, E. (2011) Weak and strong cross-section dependence and estimation of large panels. Econometrics Journal, 14: C45–C90.

Ciccone, A., Papaioannou, E. (2009) Human Capital, the Structure of Production, and Growth. The Review of Economics and Statistics, 91: 66-82.

Cohen, D., Soto, M. (2007) Growth and human capital: good data, good results. Journal of Economic Growth, 12: 51-76.

Cohen, D., Leker, L. (2014) Health and Education: Another Look with the Proper Data, CEPR Discussion Paper No. DP9940.

Currie, J., Moretti, E. (2003) Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings. Quarterly Journal of Economics, 118(4): 1495–1532.

Cutler, D.M., Lleras-Muney A. (2008) Education and health: evaluating theories and evidence. In: Making Americans Healthier: Social and Economic Policy as Health Policy, eds. Schoeni R.F., House J.S., Kaplan G.A, Pollack, H. pp. 29–60. New York: Russell Sage Found.

Deininger, K., Squire, L. (1996) New Data Set Measuring Income Inequality. World Bank Economic Review, 10: 565–591.

Eberhardt, M., Helmers, C., Strauss, H. (2013) Do spillovers matter when estimating private returns to R&D? The Review of Economics and Statistics, 95: 436–448.

Eberhardt, M., Presbitero, A.F. (2015) Public debt and growth: heterogeneity and nonlinearity. Journal of International Economics, 97: 45–58.

Eberhardt, M., Teal, F. (2013) Structural change and cross-country growth empirics. World Bank Economic Review, 27: 229–271.

Engelbrecht, H.J. (1997) International R&D spillovers, human capital and productivity in OECD economies: An empirical investigation. European Economic Review, 8: 1479-1488.

De Hoyos, R., Sarafidis V. (2006) Testing for cross-sectional dependence in panel-data models. The Stata Journal, 6(4): 482–496.

De la Fuente, A., Domenech, R. (2006) Human capital in growth regressions: how much difference does data quality make? Journal of the European Economic Association, 4: 1–36.

Feenstra, R.C., Inklaar, R., Timmer, M.P. (2015) The next generation of the Penn world table. American Economic Review, 105: 3150–3182. Retrieved at: www.ggdc.net/pwt.

Foldvari, P., van Leeuwen, B. (2011). Should Less Inequality in Education Lead to a More Equal Income Distribution? Education Economics, 19(5): 537–554.

Funke, M., Strulik, H. (2000) On endogenous growth with physical capital, human capital and product variety. European Economic Review, 44: 491–515.

Frankema, E., Bolt, J. (2006) Measuring and analyzing educational inequality: The distribution of grade enrolment rates in Latin America and the Sub-Saharan Africa. Research Memorandum No. 86, Groningen Growth and Development Centre, Groningen, Netherlands.

Gennaioli, N., La Porta, R., Lopez-de-Silanes, F., Shleifer, A. (2013) Human Capital and Regional Development, The Quarterly Journal of Economics, 128(1): 105-164.

Griffith, R., Redding, S., Reenen, J.V. (2004) Mapping the two faces of R&D: productivity growth in a panel of OECD countries. Review of Economics and Statistics, 86: 883–895.

Goll, I., Rasheed, A.A. (2005) The Relationships between Top Management Demographic Characteristics, Rational Decision Making, Environmental Munificence, and Firm Performance, Organization Studies, 26(7): 999–1023.

Green, F., Henseke, G. (2016) Should governments of OECD countries worry about graduate underemployment? Oxford Review of Economic Policy, 32: 514–537.

Hall, R.E., Jones, C.I. (1999) Why do some countries produce so much more output per worker than others? Quarterly Journal of Economics. 114: 83–116.

Gille, V. (2015) Distribution of human capital and income: An empirical study on Indian States. Journal of Macroeconomics, 43: 239-256.

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

Hanushek, E.A., Wössmann, L. (2007) The role of education quality in economic growth. Policy Research Working Paper 4122. Washington, DC : World Bank.

Hanushek, E.A., Woessman, L. (2011) How much do educational outcomes matter in OECD countries? Economic Policy, 26: 427–491.

Hooghe, M., Dassonneville, R., Marien, S. (2015) The Impact of Education on the Development of Political Trust: Results from a Five-year Panel Study among Late Adolescents and Young Adults in Belgium, Political Studies, 63: 123–41.

Hout, M. (2012) Social and Economic Returns to College Education in the United States, in K. S. Cook and D. S. Massey (eds.), Annual Review of Sociology, 38: 379–400.

Islam, N. (1995) Growth Empirics: a panel data approach. Quarterly Journal of Economics, 110: 1127–1170.

Keller, R. K. (2006) Investment in primary, secondary and higher education and the effects of growth. Contemporary Economic Policy, 24: 18–36.

Klasen, S. (2002) Does Gender Inequality in Education Reduce Growth and Development? Evidence from Cross Country Regressions, The World Bank Economic Review, 16(3): 345–373.

Knight, J.B., Sabot, R.H. (1990) Education, Productivity, and Inequality. Washington, DC: Oxford University Press.

Krueger, A., Lindahl, M. (2001) Education for growth: why and for whom? Journal of Economic Literature, 39: 1101–1136.

Lefgren L., McIntyre F. (2006) The Relationship between Women's Education and Marriage Outcomes. Journal of Labor Economics, 24(4): 787-830.

Lenkei, B., Mustafa, G., Vecchi, M. (2018) Growth in emerging economies: is there a role for education? Economic Modelling, 73: 240-253.

Levine, R., Zervos, S. (1998) Stock Markets, Banks, and Economic Growth. American Economic Review, 88: 537-558.

Lochner, L. (2011) Nonproduction Benefits of Education. Handbook of the Economics of Education, 4: 183–282.

Lochner, L.J., Moretti, E. (2004) The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-reports. American Economic Review, 94: 155–189.

López, R., Thomas, V., Wang, Y. (1998) Addressing the Education Puzzle. The Distribution of Education and Economic Reforms. World Bank Working Papers 2031.

Lucas, R.E. (1988) On the mechanics of economic development. Journal of Monetary Economics, 22: 3–42.

Mankiw, N.G., Romer, D., Weil, D.N. (1992) A contribution to the empirics of economic growth. Quarterly Journal of Economics, 107: 407–437.

Martin, T.C. (1995) Women's education and fertility: results from 26 demographic and health surveys. Studies in Family Planning, 26(4): 187–202.

Mason, G., O'Leary, B., Vecchi, M. (2012) Certified and uncertified skills and productivity growth performance: cross-country evidence at industry level. Labour Economics, 19: 351–360.

Nelson, R.R., Phelps, E.S. (1966) Investment in humans, technological diffusion, and economic growth. American Economic Review, 56: 69–75.

Pesaran, M.H., Smith, R.P. (1995) Estimating long-run relationships from dynamic heterogeneous panels. Journal of Econometrics. 68: 79–113.

Pesaran, M.H., Shin, Y., Smith, R.P. (1999) Pooled mean group estimation of dynamic heterogeneous panels. Journal of American Statistical Association, 94: 621–634.

Pistaferri, L. (2011) How much do educational outcomes matter in OECD countries? Discussion. Economic Policy, 26: 427–491.

Portela, M., Alessie, R. and Teulings, C. (2010) Measurement error in education and growth regressions. Scandinavian Journal of Economics, 112: 618-639.

Pritchett, L. (2001) Where has all the Education gone? World Bank Economic Review, 15: 367–391.

Pritchett, L. (2006) Does learning to add up add up? The returns to schooling in aggregate data. Handbook of the Economics of Education, 1: 635-695.

Ram, R., (1984) Population increase, economic growth, educational inequality, and income distribution. Journal of Development Economics, 14: 419-428.

Romer, P.M. (1990) Endogenous technological change. Journal of Political Economy, 98: 71–102.

Sander, W. (1995) Schooling and quitting smoking. Review of Economics and Statistics, 77: 191–199.

Self, S., Grabowski, R. (2004) Does education at all levels cause growth? India, a case study. Economics of Education Review, 23: 47–55.

Sianesi, B., Van Reenen, J. (2003) The returns to education: Macroeconomics. Journal of Economic Surveys, 17: 157–200.

Shapiro, J. M. (2006) Smart cities: quality of life, productivity, and the growth effects of human capital. The Review of Economics and Statistics, 88(2): 324–335.

Solow, R. (1956) A contribution to the theory of Economic Growth. Quarterly Journal of Economics, 70: 65–94.

Sunde, U., Vischer, T. (2015) Human capital and growth: Specification matters. Economica, 82: 368–390.

Squicciarini, M.P., Voigtlander, N. (2015) Human capital and industrialization: evidence from the age of enlightenment. Quarterly Journal of Economics, 130(4): 1825–83.

van Leeuwen, B., van Leeuwen-Li, J. (2015) Educational Inequality Gini Coefficient, available at: http://hdl.handle.net/10622/KORKQW

Thomas, V., Wang, Y., Fan, X. (2002) A New Dataset on Inequality in Education: Gini and Theil Indices of Schooling for 140 Countries, 1960-2000, World Bank.

Topel R. (1999) Labor Markets and Economic Growth, in eds. Ashenfelter, O. and Card, D., Handbook of Labor Economics, Amsterdam: North Holland.

Vandenbussche, J., Aghion, P., Meghir, C. (2006) Growth, Distance to Frontier and Composition of Human Capital. Journal of Economic Growth, 11: 97-127.

Appendix 1

Working out of Equation 3:

Equation 2 is is given by the following formula:

$$\ln y_{it} = \mu_{it} + \delta_{10i} \ln k_{it} + \delta_{11i} \ln k_{it-1} + \gamma_{10\mathrm{i}} \ln h_{it} + \gamma_{11\mathrm{i}} \ln h_{it-1} + \lambda \ln y_{it-1} + \varepsilon_{it}$$

If we take out $\ln y_{it-1}$ from both sides we obtain:

$$\Delta \ln y_{it} = \ln y_{it} - \ln y_{it-1} = \mu_{it} + \delta_{10i} \ln k_{it} + \delta_{11i} \ln k_{it-1} + \gamma_{10\mathrm{i}} \ln h_{it} + \gamma_{11\mathrm{i}} \ln h_{it-1} + \lambda \ln y_{it-1} - \ln y_{it-1} + \varepsilon_{it}$$

which can be re-written as:

$\Delta \ln y_{it} = \ln y_{it} - \ln y_{it-1} = \mu_{it} + \delta_{10i} \ln k_{it} + \delta_{11i} \ln k_{it-1} + \gamma_{10i} \ln h_{it} + \gamma_{11i} \ln h_{it-1} - (1-\lambda)\ln y_{it-1} + \varepsilon_{it}$

Next we both add and subtract $\delta_{10i} \ln k_{it-1}$ and $\gamma_{10i} \ln h_{it-1}$ from both sides of the equation to get:

$\Delta \ln y_{it} = \mu_{it} + \delta_{10i} \ln k_{it} - \delta_{10i} \ln k_{it-1} + \delta_{10i} \ln k_{it-1} + \delta_{11i} \ln k_{it-1} + \gamma_{10i} \ln h_{it} - \gamma_{10i} \ln h_{it-1} + \gamma_{10i} \ln h_{it-1} + \gamma_{11i} \ln h_{it-1} - (1-\lambda)\ln y_{it-1} + \varepsilon_{it}$

Simplifying:

$\Delta \ln y_{it} = \mu_{it} + \delta_{10i}\, \Delta \ln k_{it} + \delta_{10i} \ln k_{it-1} + \delta_{11i} \ln k_{it-1} + \gamma_{10i}\, \Delta \ln h_{it} + \gamma_{10i} \ln h_{it-1} + \gamma_{11i} \ln h_{it-1} - (1-\lambda)\ln y_{it-1} + \varepsilon_{it}$

$\Delta \ln y_{it} = \mu_{it} + \delta_{10i}\, \Delta \ln k_{it} + \gamma_{10i}\, \Delta \ln h_{it} + \delta_{10i} \ln k_{it-1} + \delta_{11i} \ln k_{it-1} + \gamma_{10i} \ln h_{it-1} + \gamma_{11i} \ln h_{it-1} - (1-\lambda)\ln y_{it-1} + \varepsilon_{it}$

Simplifying we get:

$\Delta \ln y_{it} = \phi_i (\ln y_{it-1} - \theta_{0i} - \theta_{1i} \ln k_{it-1} - \theta_{2i} \ln h_{it-1}) + \delta_{10i}\Delta \ln k_{it} + \delta_{20i}\Delta \ln h_{it} + \varepsilon_{it}$

Appendix 2

Construction of educational GINI coefficient

Below we include a brief description on how the educational Gini by van Leowen et al (2013) is calculated. They use the methods suggested by Thomas, Wang, and Fan (2002), Castello and Doménech (2002) and Checchi (2004). In order to convert the educational information in educational Ginis they start with

$$G^h = \frac{1}{2\overline{H}} \sum_{i=0}^{3} \sum_{j=0}^{3} \left| \hat{x}_i - \hat{x}_j \right| n_i n_j$$

Where $\overline{H}$ is average years of schooling in the population aged 15 years and over, i and j are different levels of education, n_i and n_j are the attainment per level of education, and $\hat{x}_i$ and $\hat{x}_j$ are the cumulative average years of schooling at each educational level. This equation can be rewritten as follows:

$$G^h = n_0 \frac{n_1 x_2 (n_2 + n_3) + n_3 x_3 (n_1 + n_2)}{n_1 x_1 + n_2 (x_1 + x_2) + n_3 (x_1 + x_2 + x_3)}$$

Where x stands for the average years of schooling per level of education (0= no education, 1= primary education, 2 = secondary education, and 3 is higher education) divided by the percentage population with at least that level of education attained. n_0, n_1, n_2, and n_3 are the percentages of the population with no-, primary-, secondary-, and higher education respectively. This equation gives the possibility to calculate the educational Gini for every year and country.

Appendix 3

Changes and initial levels of in human capital

Figure A1-6. Relationship between changes and initial levels of in human capital

Fig.A1

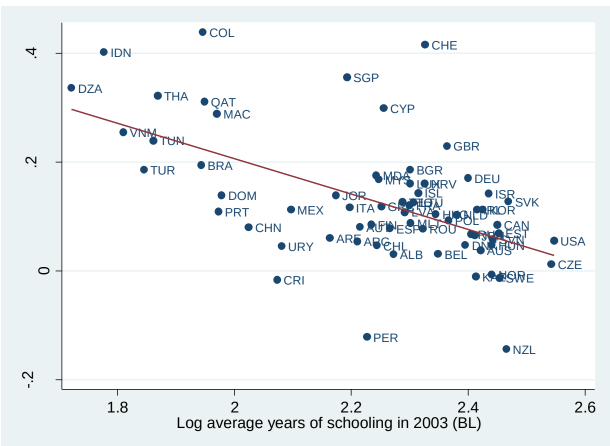


Fig.A2

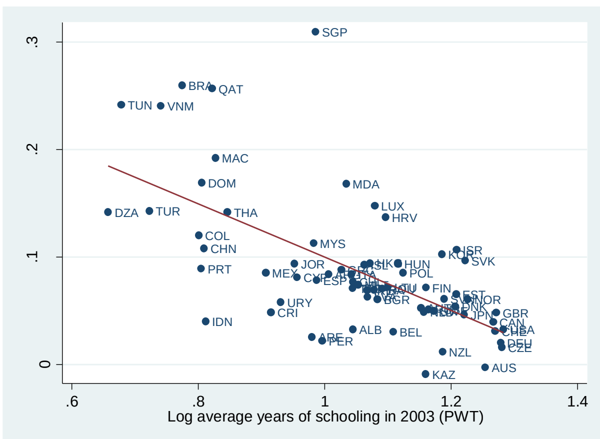


Fig.A3.

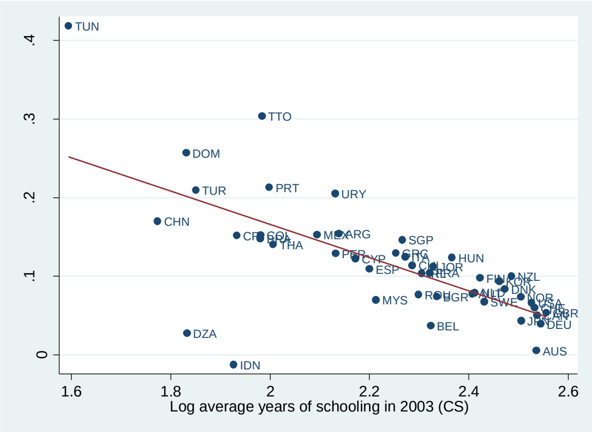
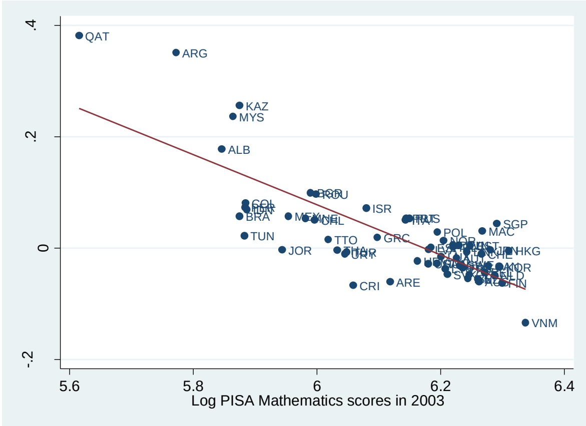
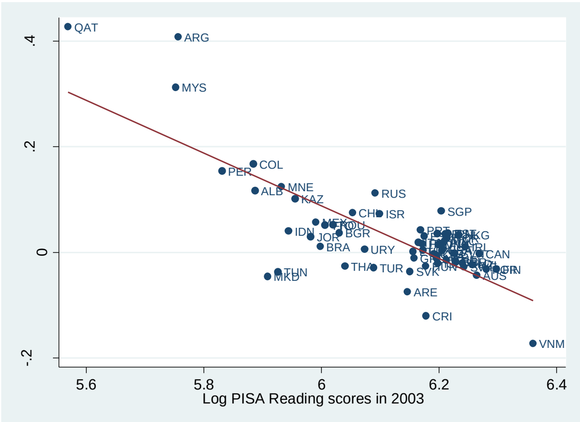


Fig.A4.



FigA5.



FigA6.

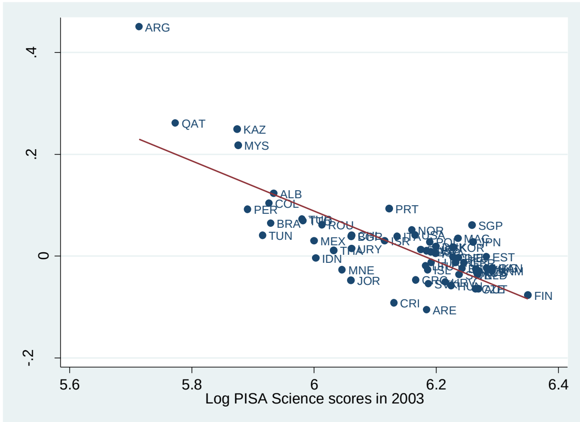
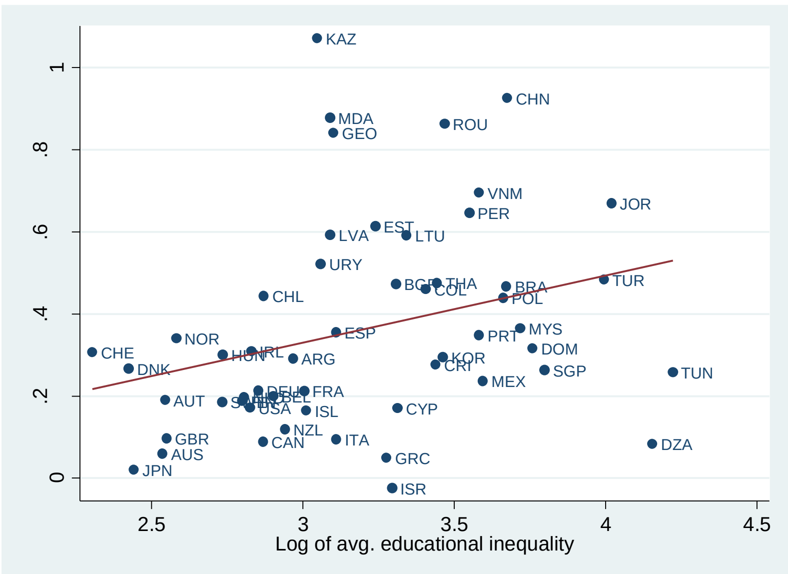


Figure A7. Relationship between change in GDP per worker

and avg. educational inequality



Appendix 4

Table A1. Growth Regressions using average schooling HC measures (short differences)

	Barro and Lee			Penn World Table			Cohen and Soto		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Δ lnh	-0.041		0.075	-0.555		-0.553	0.074		0.341
	(0.220)		(0.192)	(0.456)		(0.453)	(0.465)		(0.573)
Ln h_{t-1}		0.013	0.015		0.014	0.000		0.010	0.020
		(0.019)	(0.019)		(0.022)	(0.022)		(0.013)	(0.015)
Δ lnk	0.268**	0.266**	0.267**	0.269**	0.268**	0.269**	0.094	0.092	0.092
	(0.112)	(0.112)	(0.112)	(0.110)	(0.111)	(0.110)	(0.074)	(0.073)	(0.074)
Ln y_{t-1}	-0.020***	-0.021***	-0.022***	-0.020***	-0.021***	-0.020***	-0.022***	-0.025***	-0.026***
	(0.004)	(0.006)	(0.006)	(0.004)	(0.006)	(0.006)	(0.004)	(0.005)	(0.005)
Constant	0.100***	0.077***	0.072***	0.106***	0.091***	0.106***	0.110***	0.097***	0.077**
	(0.019)	(0.025)	(0.027)	(0.019)	(0.014)	(0.018)	(0.016)	(0.022)	(0.031)
Observations	649	649	649	649	649	649	451	451	451
R-squared	0.119	0.120	0.120	0.123	0.120	0.123	0.108	0.109	0.111

Notes: Dependent variable: annualized short differences in log GDP/worker (2003–15). Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table A2. Growth Regressions using PISA test scores HC measures (short differences)

	PISA Mathematics			PISA Reading			PISA Science		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Δ lnh	0.283		0.223	0.289		0.248	0.182		0.022
	(0.358)		(0.390)	(0.276)		(0.271)	(0.269)		(0.316)
Ln h_{t-1}		-0.016	-0.006		-0.018	-0.006		-0.020	-0.019
		(0.028)	(0.033)		(0.029)	(0.032)		(0.028)	(0.035)
Δ lnk	0.271**	0.272**	0.271**	0.275**	0.278**	0.276**	0.268**	0.274**	0.273**
	(0.112)	(0.111)	(0.113)	(0.109)	(0.108)	(0.111)	(0.111)	(0.112)	(0.115)
Ln y_{t-1}	-0.019***	-0.018***	-0.018***	-0.019***	-0.018***	-0.019***	-0.019***	-0.018***	-0.018***
	(0.004)	(0.006)	(0.006)	(0.004)	(0.006)	(0.006)	(0.004)	(0.006)	(0.006)
Constant	0.097***	0.192	0.133	0.097***	0.202	0.131	0.098***	0.215	0.209
	(0.018)	(0.153)	(0.180)	(0.016)	(0.161)	(0.177)	(0.017)	(0.156)	(0.193)
Observations	649	649	649	660	660	660	649	649	649
R-squared	0.121	0.121	0.122	0.123	0.121	0.123	0.120	0.121	0.121

Notes: Dependent variable: annualized short differences in log GDP/worker (2003–15). Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table A3. Growth Regressions using Education Inequality measures

	Educational Inequality		
	(1)	(2)	(3)
Δ lnh	0.046		0.045
	(0.080)		(0.080)
Ln h_{t-1}		-0.004	-0.004
		(0.004)	(0.004)
Δ lnk	0.088	0.088	0.089
	(0.067)	(0.067)	(0.067)
Ln y_{t-1}	-0.021***	-0.023***	-0.023***
	(0.003)	(0.003)	(0.004)
Constant	0.109***	0.126***	0.125***
	(0.013)	(0.022)	(0.022)
Observations	528	528	528
R-squared	0.111	0.112	0.113

Notes: Dependent variable: annualized short differences in log GDP/worker (2003–15).

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table A4. First Differences models: common technologies
and cross-sectional independence (1960-2015)

	OECD			Emerging Economies		
	(1)	(2)	(3)	(4)	(5)	(6)
Δlnk_{it}	0.198***	0.219***	0.193***	0.546***	0.519***	0.529***
	(0.058)	(0.058)	(0.058)	(0.060)	(0.065)	(0.058)
Δlnh_{it} (Barro-Lee)	-0.145			0.069		
	(0.164)			(0.083)		
Δlnh_{it} (Cohen-Soto)		0.339			-0.007	
		(0.364)			(0.004)	
Δlnh_{it} (PWT)			-1.021			0.067
			(0.759)			(0.600)
Constant	0.002	-0.017***	0.010	0.035***	0.004	0.005
	(0.008)	(0.005)	(0.010)	(0.013)	(0.007)	(0.008)
Observations	1,468	1,169	1,468	1,888	1,858	2,112
R-squared	0.177	0.226	0.183	0.167	0.145	0.138

Note: Dependent variable: value added per worker in log first differences. Estimation is based on a fixed effect estimator and all specifications include time and country dummies. Standard errors in brackets. * = significant at 10%, ** = significant at 5%, *** = significant at 1%.

Table A5. First Differences model: heterogeneous technologies

and cross-sectional dependence (1960-2015).

	All countries			OECD			Emerging Countries		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Δlnk_{it}	0.417***	0.406***	0.410***	0.159**	0.210***	0.157**	0.496***	0.420***	0.441***
	(0.059)	(0.051)	(0.055)	(0.074)	(0.078)	(0.072)	(0.060)	(0.072)	(0.071)
Δlnh_{it} (Barro-Lee)	0.278			0.434			0.060		
	(0.385)			(0.489)			(0.243)		
Δlnh_{it} (Cohen-Soto)		0.382			-0.327			-0.304	
		(0.327)			(0.668)			(0.247)	
Δlnh_{it} (PWT)			-0.095			0.046			-0.247
			(0.618)			(0.741)			(0.595)
Constant	0.043***	-0.006	0.060**	0.010	-0.001	0.035	0.005	0.001	0.005
	(0.016)	(0.007)	(0.026)	(0.015)	(0.005)	(0.026)	(0.008)	(0.009)	(0.009)
Observations	5,283	4,137	5,538	1,713	1,368	1,713	2,244	2,168	2,499

Note: Dependent variable: value added per worker in log first differences. All specifications are augmented with cross-section averages (not reported). Standard errors in brackets. * = significant at 10%, ** = significant at 5%, *** = significant at 1%.

Table A6. Pooled mean group estimator of the relationship between human capital and growth. OECD countries. (1960-2015)

	No correction for CSD ARDL(1,1,1)			Correction for CSD ARDL(1,1,1)			Correction for CSD ARDL(2,2,2)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\ln k_{it}$	0.712***	0.181**	0.343***	0.665***	0.197**	0.626***	0.540***	0.240***	0.556***
	(0.048)	(0.092)	(0.062)	(0.040)	(0.080)	(0.053)	(0.043)	(0.086)	(0.052)
$\ln h_{it}$ (Barro-Lee)	-0.084			0.090			0.256***		
	(0.098)			(0.085)			(0.083)		
$\ln h_{it}$ (Cohen-Soto)		1.353***			1.466***			1.216***	
		(0.267)			(0.226)			(0.217)	
$\ln h_{it}$ (PWT)			1.237***			0.623***			0.791***
			(0.199)			(0.195)			(0.182)
EC	-0.099***	-0.082***	-0.103***	-0.143***	-0.111***	-0.150***	-0.130***	-0.115***	-0.139***
	(0.019)	(0.014)	(0.019)	(0.019)	(0.017)	(0.021)	(0.019)	(0.019)	(0.024)
$\Delta\ln k_{it}$	0.011	0.153*	0.027	0.056	0.179**	0.042	0.136	0.209**	0.105
	(0.088)	(0.089)	(0.096)	(0.084)	(0.084)	(0.082)	(0.083)	(0.091)	(0.077)
$\Delta\ln h_{it}$ (Barro-Lee)	-0.392			0.244			-0.035		
	(0.396)			(0.488)			(0.383)		
$\Delta\ln h_{it}$ (Cohen-Soto)		0.206			-0.026			0.121	
		(0.644)			(0.909)			(1.011)	
$\Delta\ln h_{it}$ (PWT)			-0.185			0.616			1.355
			(0.934)			(1.033)			(1.266)
Constant	0.056***	0.008	0.107***	0.072***	-0.042***	0.031	0.101***	0.002	0.070***
	(0.008)	(0.007)	(0.015)	(0.015)	(0.013)	(0.025)	(0.019)	(0.011)	(0.026)
Observations	1,713	1,368	1,713	1,713	1,368	1,713	1,706	1,342	1,706

Notes: Dependent variable: value added per worker in log first differences. EC stands for Error Correction coefficient. Specifications (4)–(9) are augmented with cross section averages (not reported). Standard errors in brackets. * = significant at 10%, ** = significant at 5%, *** = significant at 1%.

Table A7. Pooled mean group estimator of the relationship between human capital and growth. Emerging Economies. (1960-2015)

	No correction for CSD ARDL(1,1,1)			Correction for CSD ARDL(1,1,1)			Correction for CSD ARDL(2,2,2)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
lnk_{it}	0.637***	0.662***	0.351***	0.558***	0.699***	0.566***	0.598***	0.753***	0.630***
	(0.033)	(0.034)	(0.040)	(0.040)	(0.028)	(0.026)	(0.038)	(0.025)	(0.024)
lnh_{it} (Barro-Lee)	0.218***			0.208***			0.252***		
	(0.033)			(0.042)			(0.038)		
lnh_{it} (Cohen-Soto)		0.262***			0.250***			0.266***	
		(0.026)			(0.023)			(0.023)	
lnh_{it} (PWT)			1.321***			0.532***			0.417***
			(0.100)			(0.082)			(0.069)
EC	-0.092***	-0.094***	-0.101***	-0.143***	-0.119***	-0.152***	-0.159***	-0.130***	-0.169***
	(0.011)	(0.016)	(0.013)	(0.017)	(0.019)	(0.018)	(0.017)	(0.021)	(0.020)
Δlnk_{it}	0.544***	0.388***	0.469***	0.508***	0.384***	0.436***	0.654***	0.677***	0.608***
	(0.071)	(0.086)	(0.111)	(0.077)	(0.109)	(0.103)	(0.104)	(0.148)	(0.121)
Δlnh_{it} (Barro-Lee)	-0.107			-0.031			-0.077		
	(0.218)			(0.313)			(0.303)		
Δlnh_{it} (Cohen-Soto)		-0.289			0.062			0.269	
		(0.278)			(0.314)			(0.567)	
Δlnh_{it} (PWT)			-2.107**			-1.136*			-0.720
			(0.933)			(0.672)			(0.965)
Constant	-0.002	-0.020*	0.055***	0.002	-0.079***	-0.051***	-0.036**	-0.121***	-0.091***
	(0.007)	(0.012)	(0.012)	(0.018)	(0.019)	(0.019)	(0.017)	(0.020)	(0.020)
Observations	2,244	2,168	2,499	2,244	2,168	2,499	2,229	2,126	2,479

Notes: Dependent variable: value added per worker in log first differences. EC stands for Error Correction coefficient. Specifications (4)–(9) are augmented with cross section averages (not reported). Standard errors in brackets. * = significant at 10%, ** = significant at 5%, *** = significant at 1%.

Table A8. First differences model: Educational inequality as a determinant of growth (1960-2015)

	Cross-sectional Independence		Cross-sectional Dependence	
	(1)	(2)	(3)	(4)
	OECD	Emerging	OECD	Emerging
Δlnk_{it}	0.239***	0.548***	0.207***	0.434***
	-0.051	-0.065	-0.066	-0.081
$\Delta lnGINI_{it}$	0.042	-0.005	0.054	0.032
	-0.065	-0.093	-0.099	-0.193
Constant	-0.010*	0.006	0.002	-0.003
	-0.005	-0.007	-0.003	-0.01
Observations	1,351	1,888	1,351	1,888
R-squared	0.225	0.154		
Number of id	31	42	31	42

Note: Dependent variable: value added per worker in log first differences. Estimation is based on a fixed effect estimator and all specifications include time and country dummies. Columns (1)-(3) assume common technologies and cross-sectional independence, while columns (4)-(5) assume common technologies and cross-sectional dependence. Standard errors in brackets * = significant at 10%, ** = significant at 5%, *** = significant at 1%.

Table A9. First differences model: Educational inequality as a determinant of growth, controlling for HC measures (1960-2015)

	OECD			Emerging Countries		
	(1)	(2)	(3)	(4)	(5)	(6)
	Barro Lee	Penn World Table	Cohen Soto	Barro Lee	Penn World Table	Cohen Soto
Δlnk_{it}	0.239***	0.236***	0.219***	0.566***	0.548***	0.542***
	(0.051)	(0.051)	(0.058)	(0.067)	(0.065)	(0.070)
$\Delta lnGINI_{it}$	0.043	0.032	-0.038	-0.026	-0.003	-0.023
	(0.065)	(0.054)	(0.042)	(0.072)	(0.111)	(0.091)
Δlnh_{it} (Barro-Lee)	-0.089			0.048		
	(0.159)			(0.081)		
Δlnh_{it} (PWT)		-0.951			0.049	
		(1.250)			(0.700)	
Δlnh_{it} (CS)			0.327			-0.007
			(0.353)			(0.004)
Constant	-0.009	-0.001	-0.017***	-0.003	0.005	0.004
	(0.005)	(0.013)	(0.005)	(0.010)	(0.010)	(0.007)
Observations	1,351	1,351	1,169	1,709	1,888	1,724
R-squared	0.181	0.185	0.227	0.170	0.141	0.145
Number of id	31	31	26	38	42	39

Note: Dependent variable: value added per worker in log first differences. Assuming common technologies and cross-sectional dependence Estimation is based on a fixed effect estimator and all specifications include time and country dummies. Standard errors in brackets * = significant at 10%, ** = significant at 5%, *** = significant at 1%.

Table A10. PMGE: Educational inequality as a determinant of growth

	No correction for CSD ARDL(1,1,1)		Correction for CSD ARDL(1,1,1)		Correction for CSD ARDL(2,2,2)	
	(1)	(2)	(3)	(4)	(5)	(6)
	OECD	Emerging Countries	OECD	Emerging Countries	OECD	Emerging Countries
lnk_{it}	0.578***	0.163**	0.597***	0.476***	0.567***	1.241***
	(0.043)	(0.065)	(0.056)	(0.045)	(0.058)	(0.046)
$lnGINI_{it}$	-0.316***	-0.358***	-0.06	0.018	-0.043	-0.135***
	(0.033)	(0.091)	(0.065)	(0.065)	(0.069)	(0.044)
EC	-0.090***	-0.079***	-0.086***	-0.105***	-0.103***	-0.100***
	(0.019)	(0.013)	(0.020)	(0.019)	(0.023)	(0.017)
Δlnk_{it}	0.070	0.594***	0.125*	0.571***	0.105	0.671***
	(0.072)	(0.104)	(0.069)	(0.115)	(0.073)	(0.127)
$\Delta lnGINI_{it}$	0.271**	0.465	0.182	0.103	0.198	0.303
	(0.107)	(0.358)	(0.121)	(0.280)	(0.122)	(0.278)
Constant	0.177***	0.227***	0.098***	0.023*	0.116***	-0.150***
	(0.034)	(0.034)	(0.022)	(0.012)	(0.025)	(0.028)
Obs.	1,351	1,888	1,351	1,888	1,351	1,888

Notes: Dependent variable: value added per worker in log first differences. EC stands for Error Correction coefficient. Specifications (4)–(9) are augmented with cross section averages (not reported). Standard errors in brackets. * = significant at 10%, ** = significant at 5%, *** = significant at 1%.

Chapter 2.

Time and risk preferences on parental educational expectations: Evidence from a developing country

Education in developing countries is hypothesised to be a risky investment for households, while returns are only realised after some time in the future. In this study we use data from a field experiment to analyse the relationship between parental expectations on children education attainment and parental risk and time preferences among rural households in Peru. After controlling for a set of household, parental and child characteristics, we find that our risk-aversion parameter is significantly and positively related to expectations on schooling choices. This reveals that our sample of mothers does not consider education as a risky investment in rural Peru. Instead, they perceive education as a safe option and fear of losing out on higher returns resulting from high educational attainment. Moreover, a higher degree of impatience (higher discount rate) among mothers is significantly correlated with low expectations on children's educational choices. This suggests that parents with high discount rates are expecting immediate financial gains from educational attainment of their children and therefore they may tend to underinvest in human capital given they are not patient enough to wait for its return.

1. Introduction

Despite some recent advancement in providing basic education to children in developing countries, 57 million children are still out of school, including 20 million in Latin-America (Unesco 2015). Among many determinants of children educational attainment are parental expectations about their children's educational outcomes. The role of parental expectations for children education has received considerable attention in the economics, sociology, psychology and education literature in the past decades (Seginer 1983, Goyette and Xie 1999, Glick and White 2004, Davis-Kean 2005, Yamamoto and Holloway 2010, Attanasio and Kaufmann 2015).

Parental expectations on education have substantial economic and social impact on the household. High expectations of parents lead children to set higher standards for their education and to make greater demands on themselves from an early age which in turn results in high achievement, better attendance and more positive attitudes towards school (Boocock 1972, Astone and McLanahan, 1991, Peng and Wright 1994, Reynolds 1998, Davis-Kean 2005, Yamamoto and Holloway 2010). At the same time low parental educational expectations can be transmitted through generations and they may create a vicious circle and reinforce poverty. In consequence, low parental educational expectations may reduce future academic achievement and successful labour market outcomes for children.

Past evidence in the literature shows that higher expectations translate into better academic performance and higher achieved education level (Goyette and Xie 1999, Englund et al 2004). This improved educational attainment will in turn significantly improve adult labour (Caspi et al 1998) and marriage (Ermisch et al 2006) market prospects and outcomes. Therefore, it is essential to include expectations in behavioural models that form the basis of policy decisions.

Parents form expectations[13] about future educational outcomes at early stages of their children's life (Froiland et al. 2013). Heterogeneity in educational expectations is reflected in variations of educational investments, whereas the probability of investment increases with higher expected educational returns. In turn, the heterogeneity of expectations reflects individual preferences. In this study we explore how parents' expectations vary with different levels of preferences for risk and time.

In particular, we propose that parents' expected utility derived from a child's expected level of education is a decreasing function of the riskiness of the investment and parental risk aversion[14]. Thus, more risk averse parents would require greater expected returns

---

[13] Some of the determinants that affect the formation of educational expectations include perceived returns to education and the availability of employment opportunities, observed outcomes for siblings, cultural beliefs about the connection between effort and educational success, perceived quality of education system and parental beliefs and biases.

[14] We define parental utility function using the following equation: $\mathrm{E}[U(Educ)] = \mathrm{E}(R_{Educ}) - \frac{1}{2}\mathrm{A}\sigma^2$ where, E[U(Educ)] is the expected parental utility derived from the child's expected education, $\mathrm{E}(R_{Educ})$ is the expected return for a given level of expected education, σ^2 is the variance of the investment (risk) and A is the measure of risk aversion.

to education for accepting additional risk. The literature on educational investments considers the role of risk as a result of their uncertain returns, non-marketability and difficulty to diversify accumulated human capital (Becker 1962; Levhari and Weiss 1974; Paroush 1976). In developing countries, the risks of educational investments not achieving the desired outcomes are often magnified due to poor schooling environments and uncertain qualities of public education (Glewwe and Kremer 2006), inadequate living conditions and lack of resources within the household, dangerous neighbourhoods and security issues (Tanaka and Yamano 2015). All these factors make educational investments risky choices for poor households (e.g. Glewwe and Kremer 2006; Sawada and Lokshin 2009), which can consequently shape expectations. In addition, reliance on agricultural or small family business income often results in children dropping out from schools[15] (Carneiro and Heckman 2002), meaning that those investments in the child will be sunk costs. Therefore, we claim overall that if education is perceived as a risky investment, the more risk averse the parents are, the lower expectations they will have about their children's future educational attainment.

In addition, education requires long-term commitment and years of investment are required before the returns can be realised (Becker 1964). In developing countries financial constraints can make people focus on problems in the present and neglect others in the future. Investing in the child's education can be problematic given the scarcity of resources now, compared to returns to education in the future (Mullainathan

[15] Children are further exposed to risks of unsupportive family environment such as parental neglect, harsh parenting, lack of stimulation at home or parental psychological distress.

and Shafir 2013). Since the opportunity cost of having children at school is so high, poverty regularly forces households to keep their children away from schools (Jensen and Nielsen, 1997). This is especially true for female students from less wealthy households who may have better immediate financial opportunities available through marriage. Low level of attainment is even more likely to happen if one of the parents becomes unfit to work. Since education may span over a long time period, parental patience and time preferences (the way parents discount future gains) are also contributing factors for the formation of schooling expectations.

If risk and time preferences are omitted as explanatory variables from the estimation of parental expectation models, the estimation results can be biased if these preferences are correlated with some other determinants (Dohmen et al 2010). While past studies analysed the role of risk and time preferences of parents in determining parental investments both in developed countries (e.g. Germany, Wölfel and Heineck (2012)) and developing countries (e.g. Uganda, Tanaka and Yamano (2015)), to the best of the author's knowledge no study has estimated the association between risk and time preferences of parents and parental expectations for the children in low income countries. This study is also different from previous papers in this field in the sense that there are no observations of actual parental behaviour but predictions under uncertainty about future children outcomes.

The paper is structured as follows. Section 2 presents a theoretical model showing the relationship between parental expectation and preferences. Section 3 discusses the sample used. Section 4 provides background information on the education sector in Peru.

Section 5 discusses our framework for measuring risk and time preferences and section 6 present the experimental design. Section 7 describes the empirical model used. Sections 8 and 9 present the descriptive statistics and results while Section 10 concludes.

2. Education in Peru

In order to argue whether parents consider education as a risky investment or not, we also need to understand the education system in Peru. In 1996, the government passed education reforms that extended free and compulsory school education to all students aged between 6 and 16 (primary and secondary). However, secondary school is somewhat aspirational as approximately one quarter of the relevant age group does not currently enrol in upper secondary education (UNESCO, 2013). This is especially the case in remote parts of the Andean Highlands and across the sparsely populated Amazonian rainforest in the country's interior. While public education is free, private schools operate at all levels of the education system. Schools in both public and private sectors follow the national curriculum, managed at the federal level and overseen by local education authorities, as mandated in a 2008 ministerial decree.

Education in Peru has been expanding. Since 2005, enrolment has increased to 72% for ages 3 to 5 years old; the 6 to 11-year-old age group is at 97% and the 12 to 16-year-old age group is at 91% enrolment. However, educational data tells us that 34% of children ages 5 to 17 are in the labour force.

Data from the National Institute of Statistics shows that since 1994, spending on education has quadrupled. However, Peru still only spends 3.9% of the gross domestic product on education, which is among the lowest in the world[16]. Going, forward, one of the main reasons why Peru is considering to spend more on education in the next years is the lack of quality of schools and lack of infrastructure. In 2009, the Program of International Student Assessment ranked Peru near the bottom of the 65 participating countries for reading comprehension and science, while being second to last in math. For children 7 years of age, only 13% reached required math levels and only 30% reached required reading levels. This highlights that spending money is not enough by itself as it should be spent in more productive ways.

The situation is more severe in rural areas. There may be various other reasons for drop-outs besides financial constraints including long distance from home to school, early pregnancies and lack of aspirations and interest. The lack of interest can be also attributed to overcrowding, teacher absence or other poor schools quality issues often found in rural public schools that add to the uncertainties on the returns to education and impose additional perceived costs on parents.

Vocational education is available from a variety of technological institutions; however it is mainly private. Higher education is available from both private and public universities. Since 2003, the number of students at private universities has doubled and in public universities it has increased by 12%. This substantial increase in private education shows that the more privileged students are gaining access to higher education. As public university is not completely free, it is harder for those without

[16] In 2015, in terms of % of GDP spend on education Peru ranked below other South and Central American nations including Brazil (6.2%), Mexico (5.2%) and Colombia (4.5%).

money to access it. This makes sense considering Peru has one of the highest income inequalities in the world.

On the surface, it seems as though education in Peru has somewhat improved over the last decade as literacy and enrolment rates are up as well as educational spending. However, the quality and access to education shows a different picture[17]. Therefore, increasing spending on education is a must, however investment in education needs to be conducted efficiently, considering how parents make educational decisions in low income households.

3. Framework for measuring risk and time preferences

Understanding how mothers in rural Peru form expectations should play a central role for policy-makers as individuals are likely to act upon them and they can become self-fulfilling (Armantier et al., 2012; Stinebrickner and Stinebrickner, 2012). For instance, mothers with low educational expectations are less likely to invest in their child's schooling because it is in line with their beliefs. If low expectations are associated with certain individual preferences (e.g. risk and time), they may lead to worse realised outcomes.

[17] In the 2012 Programme for International Student Assessment (PISA) test Peru ranked the worst out of 65 countries that participated in the programme. In the 2015 version it's position has slightly improved by coming 63rd out of the 69 countries taking part.

Risk preferences

We make use of the expected utility theory to explain how risk preferences may matter for expectations. It states that a rational individual maximizes the expected utility of wealth or $\sum pi * u(Wi)$ where $u(W_i)$ is the level of utility derived from wealth W_i which occurs with probability p_i for each of the *n* possible states. When the utility function is concave, the individual is said to be risk averse, preferring a sure income of W_i to a fair gamble with expected value of W_i. Using the Arrow-Pratt measure of risk aversion, $r(W)=\frac{-u''(W)}{u'(W)}$, the risk averse individual is represented by $r(W)>0$, the risk loving individual by $r(W)<0$ and the risk neutral individual by $r(W)=0$. In risk experiments, the relative risk aversion parameter, $r(M)=\frac{-M*u''(M)}{u'(M)}$, is often estimated to represent the degree of risk aversion, where *M* is the change in wealth offered in the experiment instead of the final wealth (Holt and Laury, 2002). Assuming a constant relative risk aversion (CRRA) utility function, $u(M) = \frac{M^{1-\sigma}}{1-\sigma}$, where σ indicates the curvature of the utility function and represents the risk attitude of individuals. A σ equal to 0 implies a risk neutral individual, a positive σ a risk averse, while a negative σ a risk lover person.

If parents expect rates of return to be so high that they get more than compensated for the riskiness of the investment they will have higher expectations and allocate more resources to education. We expect that the more risk averse parents are, the less investment is allocated to education if education is perceived as a risky choice. However, it is proposed that risk aversion has different impact on educational expectations that on actual observed investment behaviour. If parents perceive schooling as a safe investment that provides higher returns than any other alternatives, then the more risk adverse parents will have higher expectations for their children.

Time preferences

Present bias is the notion that individuals weight the utility from the present consumption higher than the utilities derived from consumptions at future dates, while the subjective discount rate is the rate at which individuals substitute future consumption with today's consumption. In an experimental setting, they are measured by comparing the pair-wise choice of money at two different times. The value function is $V(M_0)=\frac{1}{(1+r)^t} * V(M_t)$ where M_0 is the present value for the individual facing the payoff M_t offered at time t with r discount rate.

Our prior expectations suggest that the higher parental impatience (higher discount rate) is negatively correlated with high educational expectations for their children. This is because parents high discount rates expect immediate returns from education and likely to underinvest as they are not patient enough to wait for the returns.

4. Sample and Data

We analyse the relationship between parental expectations and risk/ time preferences using data obtained from rural households in Piura, North-Western Peru. Collection of the data took place between May and June 2016 as part of a field experiment to encourage mobile banking take-up among rural communities. Although Peru's economy has been growing rapidly in recent years, inequality still persists within the society. Between 2007 and 2012 poverty rates have dropped by around 16%, however, a quarter of the population still lives in economic hardship. The ratio of people living in poverty is even worse when comparing rural and urban areas (INEI Peru, 2013). The study was conducted by local consultancy firm Videnza and participants were all beneficiaries of JUNTOS ("together") conditional cash transfer programme. JUNTOS was launched in 2005 with two main aims: a) in the short run: providing households cash transfers thus reducing poverty and b) in the long run: reduce inter-generational transmission of poverty by investing in human capital (Perova and Vakis 2009). In order to receive a monthly lump sum cash payment of S./100 (soles)[18] beneficiaries need to comply with a number of requirements (Table 1).

Table 1. JUNTOS Conditions

For children under 5 years:	Attend regular health and nutrition controls (for periodic monitoring of height and weight, complete series of vaccinations, iron and Vitamin A supplements and anti-parasite checks)
For children 6-14 years with incomplete primary school:	School attendance at least 85% of the school year
For pregnant and breast-feeding mothers:	Attend prenatal and post-natal checks (tetanus vaccination, folic acid and iron supplements and anti-parasite checks)

Source: Perova and Vakis (2009)

[18] The median household income across Peru is S./720 monthly ($216).

As Table 1 shows all participants of the 'JUNTOS' program are required to send children to school between the ages of 6 and 14 at least 85% of the school year. If beneficiary mothers do not meet these requirements they risk losing out on higher potential income sources.

In total, survey data was collected on 1,918 households including 6,757 children. In every household survey the person filling out the questionnaire was the beneficiary mother. We focus our analysis on children aged 14 or below (N=4283) since educational expectations about maximum levels of education can be affected by actual observed outcomes at later stages of the child's schooling (e.g. end of secondary education). As such, we exclude adolescents from the analysis (Munro and Tanaka 2014).

5. Data Collection of risk aversion and time preferences

The experimental design follows that of the pairwise choice framework (Holt and Laury, 2002). Subjects were asked to complete five decision tasks. The choice options for the four time preference tasks and one risk task are presented in Appendix (Figures 1-5). For each of the four time preferences tasks subjects had to make 11 decisions, choosing between Option 1 and Option 2. The only difference between the tasks was the timing of the pay-offs. Table 1 presents the pay-off timings across the four time preferences tasks.

Table 1. Tasks 1-4. payoff timings

	Option 1	**Option 2**
Task 1	Tomorrow	day after tomorrow
Task 2	6 months	6 months and one day
Task 3	Tomorrow	6 months
Task 4	6 months	12 months

Once a subject choose Option 2 in any row, the task stops and then they automatically moved on to the next task. The purpose of these tasks is to identify the point at which the participant's preference changes from Option 1 to 2 which is indicator of the subject's discount rate. The discount rate is a parameter in an individual's utility function capturing the trade-off between the consumption at an earlier and a later date. A high discount rate represents a high degree of impatience, while the opposite is true for patient parents. For example, at Task 1 if a subject switches from Option 1 to 2 in row 5, then his or her time preference is represented by $100 \leq \frac{140}{1+r}$ or $r \leq 0.4$.

If individuals switched in the first row they were coded with the lowest discount rate (r) parameter 0, while subjects that never choose Option 2 (i.e. the highest discount rate) were coded 11. Then the four time preferences scores were added up and normalized between 0 and 1. This method allows comparability between daily and monthly discount rates.

To capture the present bias of individuals two dummy variables were created. Firstly, *presentbias_d* was coded 1 if subjects had higher switching points in Tasks 1 compared to Task 2, and 0 otherwise. In other words, present biased subjects prefer the shorter time horizon. A second present bias dummy variable (*presentbias_m*) was constructed for Tasks 3 and 4. Similarly, the variable was coded 1 had a switching point later in task 3 compared to task 4.

In the risk task subjects had to decide between Option 1, which always offered a sure payment of 100 Peruvian Soles and Option 2, a risky option where earnings were dependent on the outcome of a coin toss. Once a subject choose Option 2 in any row the game stopped there[19]. The purpose of this task was to elicit the switching point where subjects move from the sure payment to the risky option. This gives an indication of the individual subject's level of risk aversion, σ, where a large σ indicates a larger degree of risk aversion. For example, if the subject chooses Option 1 in row 3 but Option 2 in row 4, his or her risk preference can be represented as $\frac{100^{1-\sigma}}{1-\sigma} \geq 0.5^* \frac{220^{1-\sigma}}{1-\sigma} + 0.5^* \frac{0^{1-\sigma}}{1-\sigma}$ and $\frac{100^{1-\sigma}}{1-\sigma} \leq 0.5^* \frac{230^{1-\sigma}}{1-\sigma} + 0.5^* \frac{0^{1-\sigma}}{1-\sigma}$. Solving the two equations simultaneously, the interval for this individuals' risk aversion parameter is $0.132 < \sigma \leq 0.167$. The risk aversion variable is coded 0 for the lowest σ, while subjects that never chose Option 2 were coded 6. As opposed to standard laboratory experiments, the tasks on this field study were not incentivised.

6. Empirical Method

To understand the role of risk and time preferences of parents in educational expectations, we employ the following estimation model:

$$E_{ij} = \alpha_0 + \beta_1 H_i + \beta_2 C_{ij} + \beta_3 P_i + u_{ij}$$

where E_{ij} represents the measure of educational expectation of the i^{th} mother of child j; while H_i, C_{ij} and P_i represent variables at the household, child and parent level,

[19] It is important to note that our experimental tasks were not incentivised. Brañas et al. (2019) has shown that incentivising or not does not affect the elicitation of risk and time preferences using data from Nigeria and Honduras.

respectively. E_{ij} takes on the values from 0 to 10 with 0 representing low expectations (no education) while 10 the highest level (completed university education).

$H_i$includes the preference parameters, σ (risk aversion), r (discount rate) and δ (present bias). Household wealth is represented by weakly household income. The numbers of children are also included to control for the effect of competition among children in the household for the educational investment. C_{ij} includes a set of characteristics of the child j from household i such as age and gender. P_i includes parental characteristics of the household such as mother's and father's achieved highest achieved education level and mother's age. For the estimation, we cluster standard errors at the household level which is essentially the mother's identifier.

7. Descriptive Statistics

Table 2 includes the distribution of children expectations and the level of education of the parents. To elicit expectations, parents were asked 'What is the maximum educational level you think your child can reach?'. Almost 70% of the sample of children between the ages 0 and 14 are expected to attend university, 20% are expected to participate in vocational education, while 3.5% expect their child to attend below vocational level education. About 6% of our sample of children were given an uncertain answer measured by the 'Don't know' option. These expectations on children contrast with parents' own achieved education levels, which shows that in total 92% of mothers and 91% of fathers only attained secondary or lower levels of education.

Table 2. Child expectations and parental expectations distribution[20]

Child Expectations (% of total)	
Secondary or below	3.4%
Vocational	20.8%
University	69.2%
Don't know	5.7%
Mother's education level (% of total)	
Secondary or below	92.2%
Vocational	6.9%
University	0.9%
Father's education level (% of total)	
Secondary or below	91.3%
Vocational	7.7%
University	1.0%

Table 3 includes descriptive statistics of preference parameters of beneficiary mothers participating in the study and other socio-economic characteristics. It is often considered that in developing countries household heads make decisions for schooling attainment; however, this study only has preference and expectations data from the mothers. The mean maternal risk aversion parameter, σ, is 0.73 (out of 1). High degree of risk aversion among rural communities in developing countries is consistent with the literature (e.g.

[20] According to the UNESCO Institute For Statistics, in Peru the average national share of population with secondary or lower educational attainment is around 79% for both males and females. Our field study was conducted in a rural state of Peru that may explain the higher ratio of individuals with such attainment.

Barr 2003; Jacobson & Petrie 2009). The mean value of the implied discount rate parameter, *r*, used by our sample of mothers to discount future payments is 0.49 (out of 1), which suggest a degree of moderate impatience. Generally, in the literature the evidence on whether poor people are more impatient is mixed (see Cardenas and Carpenter (2006) for a review).

Table 3. Summary Statistics

	Mean	Std. Dev.
Risk and Time preferences of Mothers		
Risk aversion parameter (σ)	0.73	0.39
Discount rate parameter (r)	0.49	0.25
Present Bias (δ) – Daily	0.60	0.48
Present Bias (δ) – Monthly	0.42	0.49
Child Characteristics		
Gender (Female=1)	0.49	0.49
Age	7.63	3.97
Parental Characteristics		
Mother's age	36.32	8.28
Household Characteristics		
Weekly household income	206.33	147.75
Number of children in hh. (age 0-14)	3.29	1.69
No. of surveys	1712	
No. of children	4400	

Half of the children in the sample are female with an average age of 7.5. The average age of sampled mothers was 32.2 years old. As for the wealth indicators, the mean household weekly income in our sample is 206.33 Peruvian Soles, while the median is 180 Soles[21].

[21] Equivalent of $62 (mean) and $54 (median) per week.

8. Results

The estimation results for the determinants of child expectations are shown in Table 4. Column 1 shows the results for all children in the sample. The results show that after controlling for child, parental and household characteristics the degree of risk aversion (σ) is positively, while the daily discount rate (r) is negatively associated with expectations (Column 1). Both coefficients are statistically significant ($p<0.01$). The positive significant results between expectations and risk aversion shows that parents with high levels of risk aversion report higher expectations about their children's education as they likely perceive education for their children a less risky option. Therefore, parents consider high levels of investment in education safer. Another explanation is the fact that all subjects are participants of the JUNTOS conditional cash transfer programme. If at early stages of childhood mothers have low educational expectations and low expected returns to schooling, such expectations may affect how much parents will invest in education, starting with whether or not to send their children to school at all.

The lack of school attendance also creates the risk of losing out on monthly cash transfers which can significantly affect the budget of a low-income household. A risk averse parent makes sure that their child has adequate school attendance. Therefore, the main aim of CCT programmes of increasing attendance, may have positive spill-overs on educational expectations by raising such expectations beyond primary education. This may have important policy implications as it shows the effectiveness of conditional cash transfer programmes of improving human capital accumulation.

Columns (2) and (3) show the impact of risk preferences across age groups. Parental risk aversion of young children (0-7 years old) has relatively lower correlation with expectations compared to risk aversion for older children (8-14 year olds). An important determinant of perceived returns to education and expectations is lack of information, which increases the uncertainty of parents about the returns to schooling in the community. As the child gets older this uncertainty decreases. Thus, more information available about a child and longer exposure to observed educational outcomes at older age may explain our results on risk aversion between age groups.

Our finding on time preferences (daily discount rate) is in line with prior predictions. Impatient parents (high discount rate) tend to have lower expectations compared to patient parents (low discount rate). This is due to the fact that parents with high discount rate are expecting immediate financial gains from educational attainment of their children, and therefore they may tend to underinvest given that they are not patient enough to wait for the return. Impatient parents of younger children have lower expectations compared to impatient parents of older children (Table 3, Columns 2 and 3).

Table 4. Determinants of Child Expectations

	(1) *Expectations* *Full sample*	*(2)* *Expectations* *Age (0-7)*	*(3)* *Expectations* *Age (7-14)*
Risk (σ)	0.064***	0.054***	0.072***
	(0.016)	(0.018)	(0.019)
Time discounting (r)	-0.589***	-0.526***	-0.636***
	(0.140)	(0.176)	(0.166)
Present bias_days	-0.122*	-0.122	-0.122
	(0.069)	(0.079)	(0.084)
Present bias_months	-0.130	-0.091	-0.151
	(0.082)	(0.091)	(0.103)
Child age	0.003	0.021*	0.000

	(0.005)	(0.012)	(0.013)
Female child	-0.016	-0.030	0.007
	(0.039)	(0.055)	(0.058)
Age	0.005	0.001	0.009*
	(0.004)	(0.005)	(0.005)
Education_mother	0.059***	0.063**	0.057**
	(0.021)	(0.025)	(0.024)
Education_father	0.036*	0.062**	0.013
	(0.022)	(0.025)	(0.027)
Household income	0.001***	0.001*	0.001***
	(0.000)	(0.000)	(0.000)
No.of sibilings	-0.077***	-0.059**	-0.088***
	(0.023)	(0.029)	(0.025)
Constant	8.905***	8.800***	8.923***
	(0.220)	(0.272)	(0.295)
Observations	3,332	1,561	1,771
F	6.87	5.59	5.41
Prob>F	0.000	0.000	0.000
R-squared	0.061	0.064	0.061

Robust standard errors, in parentheses, have been clustered at the household (mother) level. *** p<0.01, ** p<0.05, * p<0.1.

Both *present bias* dummies show a negative relationship with expectations throughout. However, we only observe a weakly significant relationship (p<0.1) for the full sample at the one-day interval. The negative coefficients are in line with prior predictions. Parents affected by present bias may weight the utility from the present consumption higher than the utilities derived from consumptions at future dates. This in turn will result in lower expectations on educational attainment.

The other parental and household characteristics are significant and meeting our prior expectations as follows. Parents with higher levels of education tend to have higher expectations about their children's educational choices than those with lower levels of education. However, the coefficient on mothers' attainment have a larger impact on expectation at both the full sample and the different age groups. This can be attributed to the fact that the survey respondents were always the mothers therefore their own experience with the education system play a larger role in determining expectation for their children. The weekly household income is positively associated with educational expectations. It is likely that wealthy household tend to expect more from their children due to the available household resources. In addition, our results show that when there are higher number of siblings in the household expectations on educational attainment tends to significantly decline. This is consistent with the literature from developing countries on the propensity to concentrate resources on one child (Barrera-Osorio et al. 2011, Akresh et al. 2012). Due to the high opportunity and other costs associated with schooling and large average family sizes it is often only possible to send those children to school that are perceived to be the most likely to bring higher returns to the household. The perceived high opportunity costs are likely to drive educational expectations about the child. Another explanation to lower expectation with increasing number of siblings is that parents become more realistic with regards their expectations once they have more information available on the siblings' educational outcomes.

9. Conclusion

This paper investigates the relationship between parental expectations on child educational attainment and risk and time preferences of beneficiary mothers in rural Peru. First, after controlling for a set of household, parental and child characteristics we find that our risk-aversion parameter, sigma, is significantly and positively related to expectations on education attainment. This can be interpreted that our sample of mothers does not consider education as a risky investment in rural Peru. Instead they perceive education as a safe option and fear of losing out on higher returns resulting from high educational attainment.

Second, higher degree of impatience (higher discount rate) among parents is significantly correlated with low expectations on children attainment. Parents with high discount rates are expecting immediate financial gains from low education attainment of their children and only expect them to have higher attainment if their perceived return on education is high.

Finally, we found that expectations are not significantly affected by present bias, however the relationship is negative. This suggests that parents that weight the utility from present consumption higher than the utilities derived from future consumptions have lower expectations and may allocate less resources for future educational investments.

One of the limitations of our study is that the discount rates are highly affected by the pay-offs and the time dimensions of the task. Therefore, it is recommended that future studies explore the relationship between expectations and risk and time preferences using different parameters. The other limitation is that our subject pool only considered mothers, while it may be the father or the household head whose expectations determine future child outcomes. This can be a valid assumption in developing countries and needs further investigation.

References

Akresh, R., Bagby, E., De Walque, D., Kazianaga, H. (2012) Child ability and household human capital investment decisions in Burkina Faso. Economic Development and Cultural Change, 61(1): 157–186.

Astone, N.M., McLanahan, S. (1991) Family structure, parental practices and high school completion. American Sociological Review, 56(3): 309-320.

Attanasio, O., Kaufmann, K. (2015). Education Choices and Returns to Schooling: Mothers' and Youths' Subjective Expectations and their Role by Gender. Journal of Development Economics, 109(C): 203-216.

Barrera-Osorio, F., Bertrand, M., Linden, L., Perez, F., (2011) Improving the Design of Conditional Transfer Programs: Evidence from a Randomized Education Experiment in Colombia. American Economic Journal: Applied Economics, 3: 167–195.

Becket, G.S. (1962) Investment in human capital: a theoretical analysis. Journal of Political Economy, 70: 9-49.

Boocock, S.P. (1972) An Introduction to the Sociology of Learning. Boston: Houghton Mifflin Co.

Brañas-Garza, P., Jorrat, D., Estepa, L., Orozco, V., Rascon-Ramirez, E., (2019). To Pay or Not to Pay: Risk Preferences in the Field and Lab, mimeo.

Caspi, A., Wright, B.R.E., Moffitt, T.E., Silva, P.A. (1998) Early Failure in the Labor Market: Childhood and Adolescent Predictors of Unemployment in the Transition to Adulthood. American Sociological Review, 63(3): 424-451.

Carneiro, P, Heckman J. (2002) The evidence on credit constraints in post-secondary schooling, The Economic Journal, 112 (482): 705-734.

Davis-Kean, P.E. (2005) The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. Journal of Family Psychology, 19: 294 –304.

Dohmen, T.J., Falk, A., Huffman, D., Sunde, U. (2010) Are risk aversion and impatience related to cognitive ability? American Economic Review, 100 (3): 1238–1260.

Englund, M.M., Luckner, A.E., Whaley, G.J.L, Egeland, B. (2004) Children's Achievement in Early Elementary School: Longitudinal Effects of Parental Involvement, Expectations, and Quality of Assistance, Journal of Educational Psychology, 96(4): 723-730.

Ermisch, J., Francesconi, M., Siedler, T. (2006). Intergenerational mobility and marital sorting. The Economic Journal, 116(513): 659-679.

Froiland, J.M., Peterson, A., Davison, M.L. (2013). The long-term effects of early parent involvement and parent expectation in the USA. School Psychology International, 34: 33-50.

Glewwe, P., Kremer, M. (2006) Schools, teachers, and educational outcomes in developing countries, in E.A. Hanushek & F. Welch (ed.), Handbook of the Economics of Education, Amsterdam, 945-1017.

Glick, J.E., White, M.J. (2004) Post-secondary school participation of immigrant and native youth: the role of familial resources and educational expectations. Social Science Research, 33: 272–299.

Goyette, K., Xie, Y. (1999) Educational Expectations of Asian American Youths: Determinants and Ethnic Differences. Sociology of Education, 72: 22-36.

Hayami, Y., Godo, Y. (2004) The Three Agricultural Problems in the Disequilibrium of World Agriculture. Journal of Agriculture and Development, 3-16.

Holt, C.A., Laury, S.K. (2002) Risk Aversion and Incentive Effects. American Economic Review, 92(5): 1644-1655.

Levhari, D., Weiss, Y. (1974) The Effect of Risk on the Investment in Human Capital. American Economic Review, 64(6): 950-963.

Peng, S.S., Wright, D. (1994) Explanation of academic achievement in Asian American students. Journal of Educational Research, 87(6): 346–352.

Mullainathan, S., Shafir, E. (2013) Scarcity: Why Having Too Little Means So Much. New York: Times Books.

Munro, A., Tanaka, Y. (2014) Risky rotten kids: an experiment on risk attitudes amongst adolescents in rural Uganda. GRIPS discussion paper 14-01.

Paroush, J. (1976) The risk effect and investment in human capital. European Economic Review, 8: 339-347.

Perova, E., Vakis, R. (2009) Welfare Impacts of the 'Juntos' Program in Peru: Evidence from a Non-Experimental Evaluation. The World Bank. Technical Report.

Reynolds, A.J. (1998) Resilience among black urban youth prevalence, intervention effects, and mechanisms of influence. American Journal of Orthopsychiatry, 68(1): 84-100.

Sawada, Y., Lokshin, M. (2009) Obstacles to school progression in rural Pakistan: An analysis of gender and sibling rivalry using field survey data. Journal of Development Economics, 88(2): 335–347.

Seginer, R. (1983) Parents' educational expectations and children's academic achievements: A literature review. Merrill-Palmer Quarterly, 29: 1–23.

Tanaka, Y., Yamano, T. (2015) Risk and time Preference on schooling: Experimental evidence from a low-income country, GRIPS Discussion Paper 14-24.

Wölfel, O., Heineck, G. (2012). Parental Risk Attitudes and Children's Secondary School Track Choice. Economics of Education Review, 31(5): 727–743.

Yamamoto, Y., Holloway, S.D. (2010) Parental expectations and children's academic performance in sociocultural context. Educational Psychology Review, 22: 189-214.

Appendix

Figure 1. Task 1 – Time Preferences

What do you prefer?

	Option1	**Day (t)**	**Option 2**	**Day (t)**	**Discount Rate**
Receive…	S/. 100 tomorrow	0	S/. 100 day after tomorrow	1	$r = 0$
Receive…	S/. 100 tomorrow	0	S/. 110 day after tomorrow	1	$0 < r \leq 0.1$
Receive…	S/. 100 tomorrow	0	S/. 120 day after tomorrow	1	$0.1 < r \leq 0.2$
Receive…	S/. 100 tomorrow	0	S/. 130 day after tomorrow	1	$0.2 < r \leq 0.3$
Receive…	S/. 100 tomorrow	0	S/. 140 day after tomorrow	1	$0.3 < r \leq 0.4$
Receive…	S/. 100 tomorrow	0	S/. 150 day after tomorrow	1	$0.4 < r \leq 0.5$
Receive…	S/. 100 tomorrow	0	S/. 160 day after tomorrow	1	$0.5 < r \leq 0.6$
Receive…	S/. 100 tomorrow	0	S/. 170 day after tomorrow	1	$0.6 < r \leq 0.7$
Receive…	S/. 100 tomorrow	0	S/. 180 day after tomorrow	1	$0.7 < r \leq 0.8$
Receive…	S/. 100 tomorrow	0	S/. 190 day after tomorrow	1	$0.8 < r \leq 0.9$
Receive…	S/. 100 tomorrow	0	S/. 200 day after tomorrow	1	$0.9 < r \leq 1$
					$1 < r \leq \infty$

Figure 2. Task 2. – Time Preferences

What do you prefer?

	Option1	**Days (t)**	**Option 2**	**Days (t)**	**Discount Rate**
Receive…	S/. 100 in six months	180	S/. 100 in six months and one day	181	$r = 0$
Receive…	S/. 100 in six months	180	S/. 110 in six months and one day	181	$0 < r \leq 0.1$
Receive…	S/. 100 in six months	180	S/. 120 in six months and one day	181	$0.1 < r \leq 0.2$
Receive…	S/. 100 in six months	180	S/. 130 in six months and one day	181	$0.2 < r \leq 0.3$
Receive…	S/. 100 in six months	180	S/. 140 in six months and one day	181	$0.3 < r \leq 0.4$
Receive…	S/. 100 in six months	180	S/. 150 in six months and one day	181	$0.4 < r \leq 0.5$
Receive…	S/. 100 in six months	180	S/. 160 in six months and one day	181	$0.5 < r \leq 0.6$
Receive…	S/. 100 in six months	180	S/. 170 in six months and one day	181	$0.6 < r \leq 0.7$
Receive…	S/. 100 in six months	180	S/. 180 in six months and one day	181	$0.7 < r \leq 0.8$
Receive…	S/. 100 in six months	180	S/. 190 in six months and one day	181	$0.8 < r \leq 0.9$
Receive…	S/. 100 in six months	180	S/. 200 in six months and one day	181	$0.9 < r \leq 1$
					$1 < r \leq \infty$

Figure 3. Task 3. – Time Preferences

What do you prefer?

	Option1	**Months (t)**	**Option 2**	**Months (t)**	**Discount Rate**
Receive…	S/. 100 tomorrow	0	S/. 100 in six months	6	$r = 0$
Receive…	S/. 100 tomorrow	0	S/. 110 in six months	6	$0 < r \leq 0.016$
Receive…	S/. 100 tomorrow	0	S/. 120 in six months	6	$0.016 < r \leq 0.031$
Receive…	S/. 100 tomorrow	0	S/. 130 in six months	6	$0.031 < r \leq 0.045$
Receive…	S/. 100 tomorrow	0	S/. 140 in six months	6	$0.045 < r \leq 0.058$
Receive…	S/. 100 tomorrow	0	S/. 150 in six months	6	$0.058 < r \leq 0.070$
Receive…	S/. 100 tomorrow	0	S/. 160 in six months	6	$0.070 < r \leq 0.081$
Receive…	S/. 100 tomorrow	0	S/. 170 in six months	6	$0.081 < r \leq 0.092$
Receive…	S/. 100 tomorrow	0	S/. 180 in six months	6	$0.092 < r \leq 0.103$
Receive…	S/. 100 tomorrow	0	S/. 190 in six months	6	$0.103 < r \leq 0.113$
Receive…	S/. 100 tomorrow	0	S/. 200 in six months	6	$0.113 < r \leq 0.122$
					$0.122 < r \leq \infty$

Figure 4. Task 4. – Time Preferences

What do you prefer?

	Option1	Months (t)	Option 2	Months (t)	Discount Rate
Receive…	S/. 100 in six months	6	S/. 100 in twelve months	12	$r = 0$
Receive…	S/. 100 in six months	6	S/. 110 in twelve months	12	$0 < r \leq 0.016$
Receive…	S/. 100 in six months	6	S/. 120 in twelve months	12	$0.016 < r \leq 0.031$
Receive…	S/. 100 in six months	6	S/. 130 in twelve months	12	$0.031 < r \leq 0.045$
Receive…	S/. 100 in six months	6	S/. 140 in twelve months	12	$0.045 < r \leq 0.058$
Receive…	S/. 100 in six months	6	S/. 150 in twelve months	12	$0.058 < r \leq 0.070$
Receive…	S/. 100 in six months	6	S/. 160 in twelve months	12	$0.070 < r \leq 0.081$
Receive…	S/. 100 in six months	6	S/. 170 in twelve months	12	$0.081 < r \leq 0.092$
Receive…	S/. 100 in six months	6	S/. 180 in twelve months	12	$0.092 < r \leq 0.103$
Receive…	S/. 100 in six months	6	S/. 190 in twelve months	12	$0.103 < r \leq 0.113$
Receive…	S/. 100 in six months	6	S/. 200 in twelve months	12	$0.113 < r \leq 0.122$
					$0.122 < r \leq \infty$

Figure 5. Task 5. – Risk Preferences

What do you prefer?

Flip a coin and:

Receive…	S/. 100 sure payment	1	S/. 200 if heads and 0 if tails	2	$\sigma = 0$
Receive…	S/. 100 sure payment	1	S/. 210 if heads and 0 if tails	2	$0 < \sigma \leq 0.06$
Receive…	S/. 100 sure payment	1	S/. 220 if heads and 0 if tails	2	$0.06 < \sigma \leq 0.132$
Receive…	S/. 100 sure payment	1	S/. 230 if heads and 0 if tails	2	$0.132 < \sigma \leq 0.167$
Receive…	S/. 100 sure payment	1	S/. 240 if heads and 0 if tails	2	$0.167 < \sigma \leq 0.208$
Receive…	S/. 100 sure payment	1	S/. 250 if heads and 0 if tails	2	$0.208 < \sigma \leq 0.243$
					$0.243 < \sigma \leq \infty$

Chapter 3.

Cognitive Reflection Test: Whom, How, When

The use of the Cognitive Reflection Test as a covariate to explain behavior in Economics and Psychology experiments has significantly increased in the past few years. Experiments have shown its usefulness in predicting behavior. However, little is known about if the test is gender biased, whether incentives matter or how different implementation procedures impact outcomes. Here we report the results of a meta-study of 118 Cognitive Reflection Test studies comprising of 44,558 participants across 21 countries. We find that there is a negative correlation between being female and the overall, and individual, correct answers to CRT questions. Monetary incentives do not impact performance. Regarding implementation procedures, taking the test at the end of the experiment negatively impacts performance. Students perform better compared to non-students. We obtain mixed evidence on whether the sequence of questions matters. Finally, we find that computerized tests marginally improve results.

1. Introduction

In this meta-study we test for several of the empirical regularities regarding the Cognitive Reflection Test (Frederick, 2005) reported in several Economics and Psychology experiments. We have a heterogenous sample of studies characterized by differences in geographical location, incentives, non-student samples, lab/field based, etc. We test for whether the reported gender differences hold and whether monetary incentives significantly impact the number of correct responses in the Cognitive Reflection Test (henceforth CRT). Our meta-study also compares the CRT results for student and non-student samples of participants. We also test for whether different procedures such as the timing of the CRT, the use of computerized settings, or increased exposure to the CRT over the years has any impact on the observed results.

The CRT was first proposed by Frederick (2005) and since then has been extensively used in the Experimental Economics and Psychology literature. Frederick proposed the test based on a dual-system theory (e.g. Epstein 1994; Sloman 1996; Stanovich and West 2000; Kahneman and Frederick 2002) made up of two cognitive processes: System 1, executed quickly without much reflection and System 2, more deliberate and requiring conscious thought and effort. The questions in the CRT have an immediate (intuitive) incorrect response (System 1). However, the correct response requires some deliberation, i.e. the activation of System 2. The standard CRT test consists of the following three questions:

- *A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? (Intuitive answer 10, correct answer 5).*

- *If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? (Intuitive answer 100, correct answer 5).*
- *In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? (Intuitive answer 24, correct answer 47).*[22]

Frederick (2005) found that individuals with high CRT scores are more patient and more willing to gamble in the domain of gains. He also provided evidence that the CRT scores are highly correlated with some other tests of analytic thinking (e.g. ACT, SAT and WPT) and that males on average score higher on the test. Toplak et al. (2011) claim that the CRT can be viewed as a combination of cognitive capacity, disposition for judgement and decision making. They argue that the CRT captures important characteristics of rational thinking that are not measured in other intelligence tests. Below we discuss the results from CRT related studies.

Since Frederick (2005), several researchers have adopted the CRT as a measure of cognitive abilities and have studied its predictive power in decision making (e.g. Oechssler et al. 2009; Campitelli and Labollita 2010; Hoppe and Kusterer 2011; Besedes et al. 2012; Andersson et al. 2013; Moritz et al. 2013 etc.). Oechssler et al. (2009) investigate whether behavioral biases are related to cognitive abilities. Replicating Frederick (2005), they find that participants with low scores on the CRT are more likely to be subject to the conjunction fallacy and to conservatism in updating probabilities

[22] We will refer to the first, second and third questions as "B&B" (Bat and Ball), "Machines" and " Lillypad", respectively.

(also see Liberali et al. 2012; Alós-Ferrer and Hügelschäfer 2016). Meanwhile, Bosch-Doménech et al. (2014) find biological underpinning´s for CRT performance relating the 2D:4D ratio and performance on the CRT. They find that a lower 2D:4D ratio (reflecting a relative higher exposure to testosterone) is significantly associated with higher scores on the CRT.

The CRT has also been found to be a good predictor of the degree of strategic behavior in laboratory experiments (e.g. Brañas-Garza et al. 2012; Carpenter et al. 2013; Kiss et al. 2016 etc.). It is a useful test to measure strategic behavior as it not only captures reflective processes but also the ability to execute small computational tasks (Corgnet et al. 2015). Brañas-Garza et al. (2012) investigate the relationship between CRT outcomes and subject performance in the repeated feedback-free p-Beauty Contest Game (Nagel 1995), where a higher level of reasoning indicates better strategic behavior. They find that individuals with higher scores on the CRT choose numbers closer to the Nash equilibrium. Kiss et al. (2016) look at the effect of CRT on withdrawal decisions in an extended version of Diamond and Dybvig's (1983) bank-run game. They find that participants with higher cognitive abilities (as measured by the CRT) identify the dominant strategy when strategic uncertainty is present in the game. The above evidence indicates that the CRT could also help us in identifying strategically sophisticated subjects.

It is now well established in the Behavioural Economics and Psychological literature that subjects with better cognitive abilities are other-regarding (e.g. Ben-Ner et al. 2004; Chen et al. 2013). In recent years the link between CRT scores and social preferences has been investigated (Corgnet et al. 2015; Peysakhovic and Rand 2015; Ponti and

Rodriguez-Lara 2015; Cueva-Herrero et al 2016). Corgnet et al. (2015) find that individuals with a high CRT score are more likely to make altruistic choices in simple non-strategic decisions. Their choices increase social welfare by increasing the other person's payoff at a very low (or none) cost for the individual. On the other hand, the choices of less reflective subjects are more correlated with spiteful motives.

There is also evidence regarding the relationship between behavioral biases and cognitive reflection in the literature on behavioral finance and experimental asset markets (e.g. Cheung et al. 2014; Noussair et al. 2014; Corgnet et al. 2014; Bosch-Rosa et al. 2015; Holt et al. 2015 etc.). Corgnet et al. (2014) find that high CRT subjects earned significantly more on average than the initial value of their portfolio while low CRT subjects earned less. Interestingly, subjects with low CRT scores were net purchasers (sellers) of shares when the price was above (below) fundamental value while the opposite was true for subjects with high CRT scores. Bosch-Rosa et al. (2015) show that if subjects with only low cognitive abilities are trading in an experimental asset market it will lead to bubble formation. Further, in markets with only highly cognitive individuals assets trade close to their fundamental values. In a recent paper Holt et al. (2015) study gender differences in an experimental asset market where participants answer the standard CRT questions (with an additional mathematical question). Though they observe no gender differences in bubble formation, they find that male subjects performed better on all questions, and the difference was largest for the more mathematical (speed) question.

Another important issue is regarding gender differences. It has been shown that males consistently score significantly higher on the CRT than females (e.g. Frederick 2005;

Hoppe and Kusterer 2011; Holt et al. 2015; Cueva-Herrero et al. 2016 etc.). This agrees with the findings in the experimental literature that show that males have higher mathematical abilities and score higher than females on math tests (e.g. Benbow and Stanley 1980; Aiken 1986-1987; Benbow et al. 2000; Mau and Lynn 2010 etc.).

An important question in both Economics and Psychology is regarding the use of incentives in experiments. The effect of incentives on CRT responses has not been directly studied so far. The available evidence regarding how incentives affect outcomes is split, i.e. whether incentives matter or not is context dependent. For example, Riedel et al. (1988), Scott et al. (1988) and Duckworth et al. (2011) find a positive relationship between monetary incentives and performance levels meanwhile, others (e.g. Jenkins et al. 1998; Camerer and Hogarth 1999; Bonner and Sprinkle 2002) find evidence to the contrary. Studies that reject the impact of monetary incentives on performance outcomes argue that while it increases effort, it either doesn't improve performance at all or it only increases the performance of those who possess better cognitive abilities (Awasthi and Pratt 1990)[23].

Another issue has to do with external validity of laboratory experiments. That is, it is not clear as to how much the results from the laboratory (with university students) can be extrapolated to choices made by non-students. The evidence, again, is mixed. That is, there are mixed views on whether studies conducted with (volunteering) university

[23] The cognitive characteristic examined by Awasthi and Pratt (1990) is perceptual differentiation (PD) i.e. an individual's ability to perceptually abstract from a complex setting certain familiar concepts or relationships.

students provide reliable results (Peterson 2001; Levitt and List 2007; Falk and Heckman 2009; Falk et al. 2013; Exadaktylos et al. 2013). Common objections are that student subject pool sample sizes are small and not representative. Given our large sample we can address this issue in the context of the CRT.

We also look at the effect of positioning of the CRT compared to the main experiment. If CRT is of interest in finding covariation with decisions made in the experiment then it becomes important to understand whether timing matters. If cognitive load diminishes decision making ability then we would expect better performance the earlier is the test taken in the experiment.

Finally, we study the issue of prior experience with the CRT. This has to do with the point made by Toplak et al. (2014) where they argue that if the CRT is commonly used it is probable that individuals may have become familiarized with it. Figure 1 presents the total number of working and published papers included in our analysis over the period of 2007 to 2015. It is clear that in recent years the CRT has been increasingly used. The direct concern raised by Toplak et al (2014) is difficult to test with our data as we have no information on repeat takers of CRT. Further, note that this matter is confounded with the frequent use of the Amazon Mechanical Turk (AMT henceforth) platform for running experiments (for review on AMT see Paolacci et al. 2010; Buhrmester et al. 2011; Goodman et al. 2013). Given this study whether the year a particular CRT study was conducted and whether it was conducted on line affected test scores. This issue is also related whether different administration modes, i.e. computerized or paper-and-pencil, provide significantly different outcomes (e.g. George et al. 1992; King and Miles 1995; Cole et al. 2006 etc.). We will also be studying this with our data. Given the above we can formulate our *main* hypotheses:

Hypothesis 1: We expect that the results from our show that the CRT is gender biased. That is, males consistently perform better than females.

Hypothesis 2: Given that the evidence regarding incentives is mixed we expect no effect of incentives on CRT responses.

Hypothesis 3: A priori one would expect no difference between the student and non-student population.

Hypothesis 4: If cognitive load diminishes decision making ability then we would expect better performance the earlier is the test taken in the experiment.

Hypothesis 5: Performance in the test should improve as it is increasingly used.

The paper is organized as follows. Section 2 presents the procedures and techniques used for data collection. Section 3 provides the results. Section 4 concludes. All additional information is in the Appendix.

2. Procedures

2.1. Data collection

The information and data on the CRT were obtained through two channels. First, an e-mail inviting members of the Economic Science Association (ESA) was sent. In addition, a reminder e-mail was sent before the process was closed in June 2015. Respondents were provided with an online survey where they could input information about their study. Figure A1 (Appendix 2) presents a screen shot of the actual questionnaire that researchers were asked to fill out.

Second, we searched for research articles using the phrase "Cognitive Reflection Test" on Google Scholar. If an article was identified as one where the CRT was conducted the corresponding author was e-mailed the survey. The researchers were asked to respond to the following questions on the survey:

- Total Number of CRT participants (and the number of females among the total).
- How many of the total answered the *B&B*, *Machines*, and *Lillypad* questions correctly (and the number of females among them).
- Out of the total how many participants answered all *Three*, *Two* or *One* question(s) correctly (and the number of females among them).
- Whether the subjects received monetary incentives for correct answers.
- Whether the CRT was computerized or it was a paper and pencil test.
- The order of the CRT questions.
- Whether the CRT was conducted before, in-between or after the experiment.

We contacted 190 authors through e-mail and received information on 118 studies (62%) in total through filling out the survey (in some cases the authors had multiple studies). The corresponding authors we contacted (based on our Google Scholar search) represent roughly all the papers we could identify that used the CRT. Due to a considerable number of invitees declining to participate, our study may be hampered by self-selection bias. However, some degree of self-selection when inviting researchers to participate in a meta-analysis is almost inevitable. We still managed to obtain studies from a wide range of disciplines, both published and unpublished, and have considerable heterogeneity in our data.[24]

2.2. Sample creation

Appendix B provides a list of all research articles included in our analysis. Some research papers in our meta-analysis include two or more CRT studies. Overall our data comprises of 118 studies with 44,558 participants between the years 2007 and 2015. The articles represent a wide range of disciplines including Economics (58.1% of studies), Psychology (33.3%) and Management (2.8%) with researchers from 21 different countries[25]. The largest number of studies was conducted in the USA and Germany, 42 and 15, respectively. The study with the lowest number of observations was 40, while the study with the most had 4,312. Table 1 includes a breakdown regarding the number of observations available in each category in our sample.

The full sample of 44,558 subjects was broken down into further sub-categories. These were:

- *Female* (vs Male=0).

---

[24] However, we did not ask for papers for a specific purpose (e.g. link between the CRT and playing the Nash Equilibrium). In this sense we may expect less self-selection.

[25] These countries include (in alphabetical order): Argentina, Australia, Austria, Brazil, Canada, China, Colombia, Denmark, Finland, France, Germany, Israel, Italy, Japan, Netherlands, Slovakia, Spain, Sweden, Switzerland, UK and USA.

- Monetary *incentives* (whether the experimenter paid monetary incentives for correct answers).
- *Students* (vs Non-students=0).
- Position (whether the CRT was conducted *before*, *in-between* or *after* experiments).
- *Visibility* (the year in which the studies were conducted, see also Table 4).
- *Sequence* (the order in which the CRT questions were asked).
- *Computerized* (vs paper and pencil=0).

Table 1. Data Distribution

	Distribution (full sample)	**Distribution (regression)**
Number of studies	**118**	**118**
Total number of observations	**44,558**	**39,603**
N (Bat and Ball, Machines, Lillypad correct answers)	**41,004**	**38,031**
Bat and Ball correct	31.75%	32.24%
Machines correct	40.24%	40.84%
Lillypad correct	47.78%	48.59%
N (3,2,1 and None correct answers)	**44,558**	**39,603**
All 3 answers correct	18.17%	18.64%
Only 2 answers correct	21.12%	21.45%
Only 1 answers correct	23.18%	23.33%
None of the answers correct	37.54%	36.57%
N (gender)	**41,705**	**39,603**
Female	52.76%	52.89%
Male	47.24%	47.11%
N (monetary incentives)	**44,558**	**39,603**
Incentivized	14.67%	15.82%
Non-Incentivized	85.33%	84.18%
N (student)	**43,684**	**39,603**
Student	42.28%	41.42%
Non-Student	57.72%	58.58%
N (position of the test)	**44,558**	**39,603**
CRT took place before the experiment	37.66%	34.77%
CRT took place after the experiment	44.58%	46.46%
CRT took place in-between experiments	17.75%	18.77%
N (sequence of the questions)	**44,558**	**39,603**
Questions asked in standard sequence (B&B, Machines, Lillypad)	83.78%	84.92%
Questions asked in randomized sequence	11.64%	13.09%
Questions asked in B&B; Lilly Pad; Machines sequence	0.90%	1.01%
Questions asked in Machines; Lilly Pad; B&B sequence	2.82%	0%

Questions asked in Lilly Pad; B&B; Machines sequence	0.87%	0.97%
N (computerized or paper and pencil)	**42,797**	**39,603**
Computerized	87.91%	89.65%
Paper and Pencil	12.09%	10.35%
N (country information)	**44,217**	**39,603**
Anglo-Saxon	49.65%	46.59%
Europe	41.65%	43.70%
Rest of the world	8.70%	9.71%

2.3. Empirical strategy

We use OLS regressions to estimate the relationship between CRT outcomes and the list of variables defined earlier.[26] We use the OLS as the interpretation of its coefficients is direct. The robust standard errors are clustered around study ID's. Our meta-analysis includes 118 studies with substantial heterogeneity (e.g. paper and pencil/computerized; incentivized/non-incentivized etc.). In order to check for the robustness of our analysis we re-run our main regressions (Table 3) with six additional sub-samples (see Appendix):

- A sub-sample including female subjects only (Appendix Table ER1). In section 3.1 we analyze the impact of gender differences on CRT results.
- A sub-sample excluding studies where monetary incentives were used to reward correct answers (Appendix Table ER2). In section 3.2 we analyze the impact of monetary incentives on CRT performance.
- A sub-sample comprised of non-students (Appendix Table ER3). In section 3.3 we analyze the difference in CRT results between university student samples and samples including non-students.

[26] Other statistical models such as probit and logit provide similar results (see Appendix).

- A sub-sample excluding the studies where experiments were not conducted (Appendix Table ER4). In section 3.4 we analyze the impact of positioning of the CRT compared to the main experiment (i.e. before, in-between or after). Our general sample includes studies where the researchers did not run experiments. Having these observations in our sample could potentially lead to biased estimates. Further, by excluding these observations we can isolate the effect of these studies on the positioning of the CRT test.
- A sub-sample excluding studies where the experimenters used Amazon Mechanical Turk (Appendix Table ER5). In section 3.5 we discuss subjects' exposure to the CRT over the years. Popular online experimental platforms such as the AMT may have made the test more visible over the years. Further, the ease of access to the correct answers raises important methodological concerns[27].
- A sub-sample excluding the studies where the sequence of the questions were randomly determined (Appendix Table ER6). In section 3.6 we analyze the effect of the CRT question sequences on test outcomes. We divide our full sample between standard sequence (i.e. *B&B, Machines, Lillypad*) and other sequences. The general sample however includes studies where the sequence of questions is randomly determined. There is a 1 in 6 chance that randomization generates a standard sequence. By excluding random sequences we can isolate the effect of having standardized sequences in the other sequence sub-sample.

[27] We instantly obtained answers to all three questions through Google search.

Table 2. Mean test scores

	B&B	*Machines*	*Lillypad*	*None*	*1*	*2*	*3*
Gender							
Male	38.37%	50.43%	59.02%	27.01%	22.78%	25.00%	25.21%
Female	26.70%	32.18%	39.18%	45.09%	23.83%	18.29%	12.79%
Monetary incentives							
No monetary incentives	31.74%	39.50%	47.14%	37.82%	23.30%	20.83%	18.06%
Monetary incentives	34.76%	47.60%	55.95%	29.96%	23.53%	24.77%	21.74%
Non students vs. Students							
Non-Student	26.07%	39.21%	45.61%	40.60%	23.23%	20.05%	16.12%
Student	40.68%	43.06%	52.68%	30.88%	23.48%	23.44%	22.21%
CRT positioning							
Before	31.34%	41.94%	53.71%	33.70%	24.48%	23.04%	18.78%
In-between	32.54%	38.38%	47.30%	37.73%	23.84%	20.92%	17.51%
After	32.73%	41.08%	45.56%	38.25%	22.28%	20.48%	18.99%
Question ordering							
Non-standard order	22.60%	33.07%	34.30%	51.02%	20.31%	15.82%	12.84%
Standard order	34.01%	42.27%	51.23%	34.01%	23.87%	22.45%	19.67%
Paper and Pencil vs. Computerized							
Paper and Pencil	37.14%	36.94%	42.86%	38.66%	22.78%	19.73%	18.83%
Computerized	31.67%	41.28%	49.25%	36.33%	23.40%	21.65%	18.62%

Notes: The first three columns refer to N= 38031, while the other four columns refer to N= 39603

3. Results

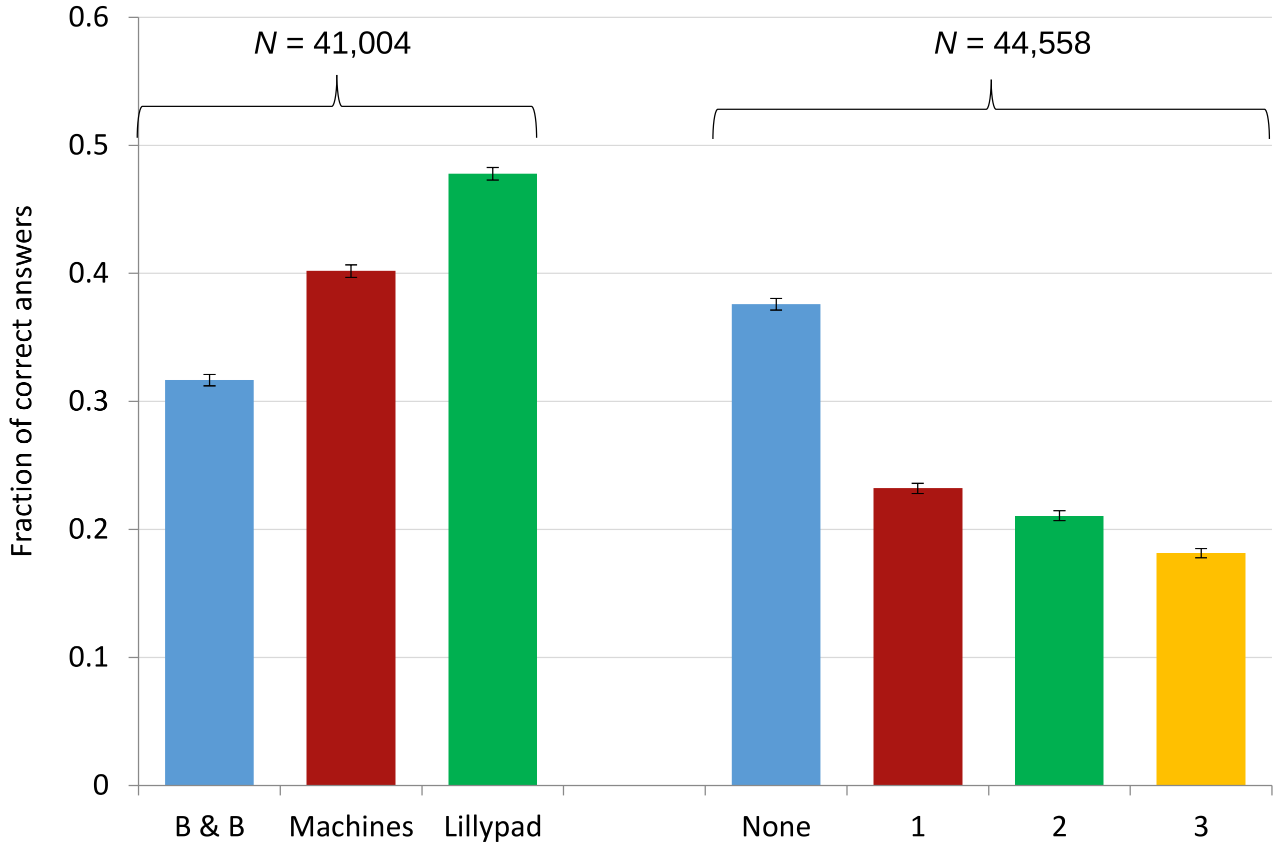
We now look at how the questions were answered both individually and overall. Figure 2 shows a summary of the results for the correct answers (by question) and for the entire test. The left side refers to the number of correct answers for each question, i.e. *B&B, Machines and Lillypad* (N = 41,004). While the *B&B* question was answered correctly by 32% in the sample, the fraction rises to 48% for the *Lillypad* question. It is hard to interpret what these proportions mean. Either the *B&B* question is more cognitively demanding for the subjects, or non-incentivized implementation (or cognitive laziness) may imply that subjects only answered the "more" intuitive questions first and did not bother answering the more cognitively difficult[28]. The two-tailed t-tests (equal/unequal variances) comparing the means of the *B&B, Machines, Lillypad* distributions reject the null hypothesis of equal means ($p<0.01$).

Looking at the total number of correct answers (right hand side, Figure 2)[29] we find that a third of the population lack reflective, or cognitive, abilities. Meanwhile, the remaining 62% have at least some, including 18% that provide all correct answers. Two-tailed t-tests (equal/unequal equal variances) comparing the distribution of the *None, 1, 2, 3* correct answers reject the null hypothesis of equal means everywhere ($p<0.01$).

Next, we study in detail the determinants of correct answers to the CRT. Moreover, in order to better understand these estimates we run a series of additional regressions in Appendix 1. Specifically, we repeat the main regression using a subsample of females only (Table ER1), a subsample of studies without monetary incentives (Table ER2), a subsample of non-students (Table ER3), a subsample of studies without economic experiments (Table ER4), a subsample excluding Amazon Mechanical Turk studies (ATM, Table ER5) and lastly a subsample of study excluding CRTs with random order (Table ER6).

[28] Note that under lack of incentives participants may choose not to answer the cognitively difficult question.
[29] Note that differences in the sample sizes are due to data availability.

Figure 2. The fraction of correct answers in the meta-study.



3.1. Gender bias

Table 2 shows that the CRT has a strong gender bias favoring males (*N* = 41,705; females 52.76%). We find that: (i) males perform better in every single question, (ii) females are more likely to answer none of the questions correctly, and (iii) males are more likely to answer all three questions correctly.

Importantly, gender differences persist in a regression (Table 3, row 1) even when we control for test characteristics (e.g. monetary incentives, computerized, student samples, positioning of the experiment etc.). Our results confirm Frederick (2005) (*N* = 3,428) who showed that males perform better in the CRT (also see Oechssler et al. 2009; Hoppe and Kusterer 2011; Holt et al. 2015; Cueva-Herrero et al. 2016 etc.).

We replicated the regressions in studies without incentives (Tables ER2), in a subsample of non-students (ER3), in studies without experiments (ER4), in Non-AMT studies (ER5), and in studies without randomly sorted questions (ER6). Our replications show that the gender bias remains negative and statistically significant ($p<0.01$) throughout. Hence, we find that all previous results hold.

In sum, gender has an important impact on CRT performance and if used as a sorting criteria may bias the distribution of participants. This gives us our first result,

Result 1: Our results strongly support the Hypothesis 1. CRT responses have a strong gender bias. The proportion of males is increasing with increase in the score.

This is a useful find as knowing that the CRT has a strong male bias can be important for sample building. For instance, say that we would like to select subjects with certain characteristics from the sample. Our study suggests that using the 3-correct-answers criteria will give us twice as many males than females. This implies that we not only select highly cognitive individuals, but also that the sample is strongly biased towards males.

Bosch-Rosa et al. (2015), for example, divide their subject pool between individuals with low and high cognitive abilities based on the CRT results in order to perform a later task. Our results suggest that their findings might be partly driven by gender effects. A similar problem arises in Brañas-Garza et al. (2012) where they find that high CRT scorers are more likely to play according to the Nash Equilibrium in the Beauty Contest

Game. This may again be due to the higher proportion of males rather than just an overall effect of high CRT scorers.

3.2. Incentives

The effect of financial incentives on human behavior has been a long debated issue in Economics and Psychology (for a review see Camerer and Hogarth 1999). The dominant argument in the experimental methodology is that incentives are important for profit maximizing individuals. In our case this would imply that the number of correct answers would improve under monetary incentives (14.67% of our sample).

Table 3. Regression analysis

	(1) *B&B*	(2) *Machines*	(3) *Lillypad*	(4) *None*	(5) *1*	(6) *2*	(7) *3*
(1) female	-0.113***	-0.177***	-0.197***	0.179***	0.009	-0.066***	-0.121***
	(0.011)	(0.010)	(0.010)	(0.010)	(0.006)	(0.007)	(0.008)
(2) monetary incentives	-0.026	0.003	0.040	-0.005	-0.002	0.000	0.008
	(0.046)	(0.048)	(0.049)	(0.045)	(0.016)	(0.017)	(0.040)
(3) student	0.138***	-0.002	0.067*	-0.089**	0.011	0.030**	0.047*
	(0.035)	(0.025)	(0.039)	(0.034)	(0.008)	(0.013)	(0.024)
(4a) in-between experiments	-0.046	-0.007	-0.090*	0.059	0.002	-0.017	-0.043
	(0.045)	(0.035)	(0.049)	(0.040)	(0.013)	(0.014)	(0.033)
(4b) after the experiment	-0.032	-0.009	-0.093**	0.060*	-0.008	-0.026**	-0.026
	(0.037)	(0.030)	(0.038)	(0.035)	(0.009)	(0.012)	(0.026)
(5) visibility	0.008	0.016***	0.005	-0.005	-0.007***	0.002	0.010**
	(0.006)	(0.006)	(0.006)	(0.005)	(0.002)	(0.002)	(0.005)
(6) standard sequence	0.103**	0.102***	0.148***	-0.142***	0.012	0.050***	0.080***
	(0.040)	(0.034)	(0.043)	(0.040)	(0.012)	(0.015)	(0.031)
(7) computerized	0.033	0.085*	0.108**	-0.095**	0.013	0.050**	0.032
	(0.038)	(0.048)	(0.051)	(0.045)	(0.012)	(0.020)	(0.032)
Constant	0.184**	0.270***	0.285***	0.533***	0.241***	0.156***	0.070
	(0.072)	(0.073)	(0.074)	(0.074)	(0.022)	(0.030)	(0.056)
N	38031	38031	38031	39603	39603	39603	39603
R^2	0.045	0.052	0.071	0.067	0.003	0.015	0.038

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, *p<0.1. The regressions also control for country by using two dummies: europe and anglo-saxon.

The regression analysis (row 2, Table 3) shows that the variable *monetary incentives* is not statistically significant at any of the common significance levels. This implies that paying subject for correct answers on the CRT does not increase performance levels.

Our robustness checks show that effect of incentives are only marginally significant for the female subsample (Tables ER1), for non-students (ER3) and in studies without experiments (ER4); while the lack of-effect of *monetary incentives* remains persistent throughout in Non-AMT studies (ER5) and in studies without randomly sorted questions (ER6). Below we present our result,

Result 2: Our results support Hypothesis 2. We find that overall incentives have no impact on CRT performance.

Note that the role of incentives and the degree of cognition can also be important. For example, Awasthi and Pratt (1990) find that the effectiveness of monetary incentives depends on the cognitive skill of the decision maker. In their study monetary incentives were associated with higher performance only for higher cognition individuals. We cannot comment on whether there is a relation between cognition and incentives. One may also argue that the test was a marginal part of a larger study and payments were not salient (Gneezy and Rustichini 2000). Finally, we should point out that we lack specific details on how incentives were implemented and their magnitude.

3.3. Students vs. non-students

Economics experiments have been traditionally run with university students. This has raised an obvious question about external validity of experimental data (see Levitt and List 2007; Falk and Heckman 2009; Exadaktylos et al. 2013). Our sample includes studies that were conducted with, and without, university students (42.28% of all observations). This allows us to check for whether there are population differences in the CRT.

Overall we find that the student population performs better than non-students. We find that students score significantly better in the *B&B* and, only slightly better in the *Machines* and *Lillypad* question (Table 2). The results in Table 2 also show that university students are less likely to have all three questions answered incorrectly, while at the same time they are more likely to give two and three correct answers. Below we summarize our results,

Table 3 (row 3) confirms the findings in Table 2. The *student* coefficient is statistically significant for the *B&B* $(p<0.01)$ and *Lillypad* $(p<0.1)$ questions implying that students are more likely to give correct answers to these two questions. In contrast, the coefficient for zero correct answers is negative and statistically significant at the *5%* level. This implies that non-students on average are more likely to obtain all incorrect answers relative to students. Furthermore, students are more likely to have two $(p<0.05)$ and all three $(p<0.1)$ answers given correctly. Results on the high performance of students compared to non-students are likely to be derived from higher cognitive ability of students compared to average population (e.g. Pennycook et al. 2012).

Our robustness checks show that these effects have similar signs but less statistical power for the female subsample (Table ER1), subsample without monetary incentives (ER3), in studies without experiments (ER4) and in studies without randomly sorted questions (ER6), however similar significance levels for the subsample using only Non-

AMT studies (ER5). In sum, our results allow us to state that one can expect the average CRT scores to be higher when using *student* samples.

Result 3: Hypothesis 3 is not supported. We find that students perform significantly better than non-students.

3.4 When?

If CRT is of interest in finding covariation with decisions made in the experiment then it becomes important to understand whether timing matters. In our sample the proportion of studies where the test was conducted *before*, *in-between* or *after* the experiment is 37.66%, 17.75% and 44.58%, respectively. A priori one would expect no differences. However, there are reasons why the timing may be important. The first is cognitive load. If students perform cognitively difficult tasks in the experiments then a later CRT would imply higher cognitive load and hence may affect CRT response rates. The second argument could be related to glucose depletion. It has been shown that brain activity is reliant on blood glucose levels as it affects the firing of neurons (Weiss 1986). Experimental tasks almost always require some form of cognition (reading instructions, answering questionnaires, quizzes etc.) and it would be reasonable to assume that glucose levels would be lower towards the end of the experiment. This would then consequently imply that if the CRT is conducted at the end of the experimental then performance on the CRT should be negatively affected[30]. Below we present our results,

The main message from our analysis is that CRT efficiency declines the later it is conducted. One sees that there are some differences in CRT performance depending upon whether it was conducted *before*, *in-between* or *after* the experiment (see rows 4a and 4b in Table 3). Conducting it *in-between* or *after* has a negative and statistically significant effect on the *Lillypad* question ($p<0.1$ and $p<0.05$, respectively) (rows 4a and 4b, Table 3). In addition, conducting it *after* is more likely to result in *None* ($p<0.1$) and less likely to have exactly two questions answered correctly ($p<0.05$). It is important to note that the *after-the-experiment* coefficient remains negative throughout (row 4b, Table 3). This suggests that conducting the CRTs *after the experiments* can potentially impact outcomes negatively.

Note, however, that prior data includes studies where no experiments were conducted. We conducted further analysis by removing these studies from the sample. This gives us even stronger results (Table ER4, rows 4a and 4b). Now it is even less likely that subjects are to answer the *B&B* and *Lillypad* questions correctly if CRTs conducted *in-between* or *after the experiments*. This negative effect is lower for *in-between* experiments ($p<0.05$) and stronger for *after* the experiments ($p<0.01$) variables. The stronger negative effect for the variable *after* is coherent with the argument that glucose levels are being depleted as subjects are progressing through the experiment. Similarly, we observe that

[30] People performing worse on the CRT at the end of the experiment can be also confounded by the fact that they may have had less time then. For example, the experiment running late. In addition, experiments that measure CRT at the end may be different and typically longer.

subjects are less likely to answer all three questions correctly both *in-between* and *after experiments* (both $p<0.05$) and more likely to have *None* (both $p<0.05$) (rows 4a and 4b, Table ER4).[31]

Result 4: Our results support Hypothesis 4. Performance in the CRT improves the earlier the test administered in the experiment. The results are stronger excluding CRT implementation without an experiment.

Whether it is cognitive load or glucose depletion it is important to know that performance in the test gets worse the later it is conducted in the experiment. It is known that that glucose levels (in the brain) play an important role in cognition. Effortful, controlled or executive processes and tasks (e.g. experiments) require more glucose than simpler, less effortful or automatic processes. When glucose levels are low, cerebral functioning is disrupted, producing numerous cognitive and behavioral deficits (Gailliot and Baumeister 2007). In sum, our results show that conducting the CRT after the experiment can have a negative effect on performance on the CRT[32].

[31] The robustness checks (Tables A3, A4, A6, A8) report similar results with varying degrees of significance.

[32] Poor performance may also be due to lack of effort, or leaving problems blank. We lack data on whether an incorrect answer is wrong or empty. We would like to thank Shane Frederick for pointing this out.

3.5. Exposure to the CRT over the years (visibility)

Toplak et al. (2014) argue that the test in its original form is becoming increasingly popular and is perhaps losing its efficacy. This argument has validity if the student pool remains the same, or same subjects take the test on more than one occasion over their University life. The critique is of concern given the increased implementation of the CRT and if we are to believe in its predictive power. This issue is also related with the fact that some studies are conducted on-line. Answers to the CRT are easily available online and this sheds doubt on its efficacy using online studies. We investigate these issues below.

In our regressions we used the variable *visibility* to describe the effect of exposure to the CRT over the years. The variable was generated by assigning the value 0 for studies conducted in 2005, 1 for 2006 and so on.

In Table 3 (row 5) the variable *visibility* shows that the number of years of exposure has a positive impact on obtaining all three correct answers ($p<0.05$). *Visibility* negatively affects subjects answering only one question correctly ($p<0.01$), the coefficients on two and three correct answers turn positive but non-significant. No effect is found for *None* ($p>0.1$), i.e. exposure to the test is not decreasing the number of participants giving zero correct answers. In addition, we find that subjects are more likely to answer the *Machines* question correctly ($p<0.01$). Overall, some support (row 5, Table 3) is lent to the argument that years of exposure positively affect test outcomes. This effect, however, does not seem to be too large or persistent. The robustness checks provide similar weak findings for the exposure conjecture for the females only subsample (Table ER1), for non-students (Table ER3), studies without economic experiments (Table ER4) and for studies excluding CRTs with random order (Table ER6). However, results show higher statistical power for the subsample of studies without monetary incentives (Table ER2).

The earlier results could be confounded by the presence of AMT studies in our sample. AMT studies have the potential problem of immediate internet access enabling easily access to the answers for the standard questions. Our results change when AMT studies are excluded from the sample (row 5, Table ER5). We find that the previously significant effects on *visibility* are substantially weaker. All in all, we cannot observe a clear link between length of exposure and obtaining correct answers in the CRT. This gives us the result below,

Result 5: Hypothesis 5 is not supported by our results. Excluding the AMT studies we find little support for the hypothesis that performance on the test improves as its exposure increases.

Besides our main hypotheses we also check for how using the standard test sequence, or hand-run vs computerized implementation impacts CRT performance. Below we present these results.

3.6. The sequence of questions

The most commonly used sequence for the CRT is the one originally proposed by Frederick (2005), i.e. *B&B*, *Machines*, and *Lillypad.* A large proportion of our sample, *83.78%,* corresponds to this. It is thus reasonable to see whether the standard implementation of the sequence affects final outcomes. Looking at all the studies (standardized and random sequences) we see that subjects score better on the CRT when the questions are presented in the standard order (Table 3). However, excluding studies with randomized and other forms of implementation we find that the effect of standardized implementation on CRT responses is marginal. We thus find no clear evidence on the effect of implementing the standard sequence upon outcomes.

Looking at both the standard and randomized studies we find that the coefficient of *standard* sequence (row 6 - Table 3) is significant for the *B&B* $(p<0.05)$, *Machines* $(p<0.01)$ and *Lillypad* $(p<0.01)$ questions. Further, the likelihood of *None* is much higher when the questions are not asked in the standard order $(p<0.01)$. Likewise, subjects are more likely to answer two $(p<0.01)$ or three $(p<0.01)$ questions for the standard implementation. This result is, however, not robust.

Note that the randomized sequences can also include questions asked in a standard way with probability 1 in 6. [33] Controlling for 'other sequences' and excluding studies where the order of the questions was randomized (11.64% of all of our observations) we find that the effect is marginal. In Table ER5 we replicated the main regressions excluding the studies with random sequences. The effect of standardized sequence on correct CRT responses is now marginal (Row 5, Table ER6). We cannot thus conclude that the standardized sequence would bias responses in the CRT.

[33] If we consider that 1/6 of the randomized sample use standard sequence (roughly 2% of the sample) then we have that 85.7% of the sample uses the standard sequence and 14.3% non-standard (including 5/6 of the random).

3.7 Hand run vs. computerized?

Next we explore whether different administration modes effect performance on the test. In this case one would not expect that either method of implementation expects outcomes as it is a problem solving task not involving specific decisions (as occurs in most experiments). However, it is still interesting to study whether different forms of implementation affect final outcomes.

We find (Table 3, row 7) that the dummy variable for *computerized* is only weakly significant. We do observe that subjects using computers are less likely *($p<0.05$)* to fail all three questions and more likely to have two correct answers *($p<0.05$)*. Further, computerized implementation favors performance in the *Machines ($p<0.1$)* and *Lillypad* questions *($p<0.05$)*, however, we do not observe significant effects on the *B&B* question *($p>0.1$)*. We find this puzzling since one would expect that using paper and pencil would be more conducive to obtaining correct answers.

Our robustness check show similar but slightly more significant results for the subsample of females (Table ER1) and the studies without monetary incentives (Table ER2) and the subsample of studies excluding CRTs with random order (Table ER4), however, the results show less statistical power for the subsample of non-students (Table ER3). Finally, the subsample of studies without economic experiments (Table ER4) and the subsample excluding Amazon Mechanical Turk studies (ATM, Table ER5) produce similar results. Note, however, we do not have information on whether participants could work out solutions on paper while responding to the computerized questions. Summarizing, we find that running the CRT on computers as compared to paper and pencil results in weakly significant positive effects on test scores.

4. Discussion

The CRT has become increasingly popular in predicting behavior in Economics and Psychology experiments. However, there is no consensus on how the vastly different implementation procedures used, i.e. being incentivized or not, administered by paper-and-pencil/computers/AMT, before/in-the-middle/after-an-experiment, etc. impact performance on the CRT. We only know from Frederick (2005) that the test has a strong (male) gender bias. The purpose of this study is to provide the first extensive look at how different implementation procedures for CRT may impact performance on the test. In the end if the CRT is useful for its predictive power then knowing whether any small variation in implementation procedures can affect outcomes is important.

This study conducts a meta-survey of the methods employed in 118 studies ($N = 44{,}558$) that use CRT. Our main result reaffirms and provides additional findings regarding the gender bias result first reported in Frederick (2005). We find that males perform notably better in this test. This observation is important if, say, one is interested in constructing samples based on cognitive ability. This could lead to strong (gender) sample imbalance. For instance, if one uses three correct answers as a selection criteria then the sample is disproportionately biased towards males. Our second interesting finding is that we find no statistical evidence to support the argument that *monetary incentives* may play an important role in improving CRT performance. Albeit limited (as we lack data on the amount, or how, subjects were paid), this result is important as it tells us that incentives may not be strongly relevant for the implementation of the CRT. Regarding comparing student vs non-student populations we find that *students* are more likely to answer all three questions correctly compared to non-students, and less likely to have zero correct answers. Again this tells us that the predictive power of the CRT may be affected by population differences.

We also find that conducting the CRT *after the experiments* negatively effects test outcomes. Conducting the test later decreases the probability of obtaining correct answers; meanwhile, the probability of obtaining *None* is increased. This result is interesting as it points towards the fact that increased cognitive load could be an important determinant of performance in the CRT. Another interpretation of this result could be that it provides indirect support to the argument that glucose is important in cognitive tasks and cognition declines with time and effort. This is important as after removing studies where the researchers did not run experiments from the data we find even more significant results. We test for the year effect (*visibility*) and find no clear evidence that exposure positively affects tests results.

Comparing test scores for hand-run vs. *computerized* tests we found a weakly positively significant effect of computerized implementation of the test. It is important to point out that we do not collect individual CRT scores but session information about CRT score distribution and do not control for individual characteristics such as cognitive ability, for example measured by IQ. This makes the analysis of individual characteristics challenging. Finally, we should add that, as is common with studies of this nature, a comprehensive list of data was not available. We lacked information about particular details (such as length of experiment, size of incentives, etc.) of each experiment in our meta-study. Knowing these details would have aided the interpretation of our results.

References

Aiken, L. (1986-1987) Sex differences in mathematical ability: A review of the literature. Educational Research Quarterly, 10: 25-35.

Alós-Ferrer, C., & Hügelschäfer, S. (2016) Faith in intuition and cognitive reflection. Journal of Behavioral and Experimental Economics, 64: 61–70.

Andersson, O., Tyran, J.R., Wengström, E., & Holm, H.J. (2013) Risk aversion relates to cognitive ability: Fact or fiction? IFN Working Paper No. 964.

Awasthi, V., Pratt, J. (1990) The effects of monetary incentives on effort and decision performance: The role of cognitive characteristics. The Accounting Review, 65(4): 797-811.

Benbow, C.P., Stanley, J.C. (1980) Sex differences in mathematical ability: Fact or artifact? Science, 210(4475): 1262–264.

Benbow, C. P., Lubinski, D., Shea, D. L., Eftekhari-Sanjani, H. (2000) Sex differences in mathematical reasoning ability: Their status 20 years later. Psychological Science, 11: 474-480.

Ben-Ner, A., Kong, F., Putterman, L. (2004) Share and share alike? Gender-pairing, personality, and cognitive ability as determinants of giving. Journal of Economic Psychology, 25(5): 581-589.

Besedes, T., Deck, C., Sarangi, S., & Shor, M. (2012) Decision-making strategies and performance among seniors. Journal of Economic Behavior and Organization, 81(2): 524-533.

Bonner, S.E., Sprinkle, G.B. (2002) The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. Accounting, Organizations and Society, 27: 303–345.

Bosch-Domènech, A., Brañas-Garza P., Espín A.M. (2014) Can exposure to prenatal sex hormones (2D:4D) predict cognitive reflection? Psychoneuroendocrinology, 43: 1-10.

Bosch-Rosa, C., Meissner, T., & Bosch-Domenech, A. (2015) Cognitive bubbles. SFB 649 Discussion Paper 2015-006.

Brañas-Garza, P., García-Muñoz, T., González, R.H. (2012) Cognitive effort in the beauty contest game. Journal of Economic Behavior & Organization, 83(2): 254–260.

Buhrmester, M., Kwang, T., & Gosling, S.D. (2011) Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? Perspectives on Psychological Science, 6(1): 3-5.

Camerer, C. F., Hogarth, R.M. (1999) The effects of financial incentives in experiments: A review and capital-labor-production framework. Journal of Risk and Uncertainty, 19(1): 7-42.

Campitelli, G., Labollita, M. (2010) Correlations of cognitive reflection with judgments and choices. Judgment and Decision Making, 5(3): 182-191.

Carpenter, J., Graham, M., Wolf, J. (2013) Cognitive ability and strategic sophistication. Games and Economic Behavior, 80(1): 115–130.

Chen, C.C., Chiu, I.M., Smith, J., Yamada, T. (2013) Too smart to be selfish? Measures of cognitive ability, social preferences, and consistency. Journal of Economic Behavior & Organization, 90(0): 112–122.

Cheung, S. L., Hedegaard, M., Palan, S. (2014) To see is to believe. Common expectations in experimental asset markets. European Economic Review, 66: 84–96.

Cole, M.S., Bedeian, A.G., Field, H.S. (2006) The measurement equivalence of web-based and paper-and-pencil measures of transformational leadership: A multinational test. Organizational Research Methods, 9(3): 339-368.

Corgnet, B., Espin, A.M., Hernan-Gonzalez, R., Kujal, P., Rassenti, S. (2016) To trust, or not to trust: Cognitive reflection in trust games. Journal of Behavioral & Experimental Economics, 64: 20-27.

Corgnet, B., Espín, A.M., Hernán-González, R. (2015) The cognitive basis of social behavior: cognitive reflection overrides antisocial but not always prosocial motives. Frontiers in Behavioral Neuroscience, 9(287).

Corgnet, B., Hernan-Gonzalez, R., Kujal, P., Porter, D. (2014) The effect of earned versus house money on price bubble formation in experimental asset markets, Review of Finance, 1-34, [doi: 10.1093/rof/rfu031]

Cueva-Herrero, C., Iturbe-Ormaetxe, I., Mata-Prez, E., Ponti, G., Sartarelli, M., Yu, H., & Zhukova, V. (2016) Cognitive (ir)reflection: New experimental evidence. Journal of Behavioral and Experimental Economics, 64: 81-93.

Diamond, D.W., & Dybvig, P.H. (1983) Bank runs, deposit insurance, and liquidity. Journal of Political Economy, 91(3): 401-419.

Duckworth, A.L., Quinn, P.D., Lynam, D.R., Loeber, R., & Stouthamer-Loeber, M. (2011) Role of test motivation in intelligence testing. Proceedings of the National Academy of Sciences, 108: 7716–7720.

Epstein, S. (1994) Integration of the cognitive and the psychodynamic unconscious. American Psychologist, 49: 709–724.

Exadaktylos, F., Espín, A.M., Branas-Garza, P. (2013) Experimental subjects are not different. Scientific Reports, 3(1213): 1-6.

Falk, A., Heckman, J. (2009) Lab experiments are a major source of knowledge in the social sciences. Science, 326: 535–538.

Falk, A., Meier, S., Zehnder, C. (2013) Do lab experiments misrepresent social preferences? The case of self-selected student samples. Journal of the European Economic Association, 11(4): 839–852.

Frederick, S. (2005) Cognitive reflection and decision making. Journal of Economic Perspectives, 19(4): 25-42.

Gailliot, M.T., Baumeister, R.F. (2007) The physiology of willpower: Linking blood glucose to self-control. Personality and Social Psychology Review, 11(4): 303-327.

George, C.E., Lankford, J.S., Wilson, S.E. (1992) The effects of computerised versus paper-and- pencil administration on measures of negative affect. Computers in Human Behavior, 8(2-3): 203-209.

Gneezy, U., Rustichini, A. (2000a) Pay enough or don't pay at all. The Quarterly Journal of Economics, 115(3): 791–810.

Goodman, J. K., Cryder, C.E., Cheema, A.A. (2013): Data collection in a flat world: strengths and weaknesses of Mechanical Turk samples. Journal of Behavioral Decision Making, 26: 213-224.

Holt, C.A, Porzio, M., Song, M.Y. (2015) Price bubbles, expectations, and gender in asset markets: An experiment. University of Virginia working paper.

Hoppe, E. I., Kusterer, D.J. (2011) Behavioral biases and cognitive reflection. Economics Letters, 110: 97–100.

Jenkins, G. D., Mitra, A., Gupta, N., Shaw, J.D. (1998) Are financial incentives related to performance? A meta-analytic review of empirical research. Journal of Applied Psychology, 83(5): 777-787.

Kahneman, D., Frederick, S. (2002) Representativeness revisited: At- tribute substitution in intuitive judgment. In T. Gilovich, D. Griffin and D. Kahneman (Eds.), Heuristics and biases: The psychology of intuitive judgment, 49-81, New York: Cambridge University Press.

King, W.C., Miles, E.W. (1995) A quasi-experimental assessment of the effect of computerizing noncognitive paper-and-pencil measurements: A test of measurement equivalence. Journal of Applied Psychology, 80: 643-651.

Kiss, H.J., Rodriguez-Lara, I., Rosa-García, A. (2016) Think twice before running! Bank runs and cognitive abilities. Journal of Behavioral and Experimental Economics, 64: 12-19.

Levitt, S.D., List, J.A. (2007) What do laboratory experiments measuring social preferences reveal about the real world? Journal of Economic Perspectives, 21(2): 153–174.

Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., Pardo, S. T. (2012) Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. Journal of Behavioral Decision Making, 25(4): 361–381.

Mau, W.C., Lynn, R. (2010) Gender differences in homework and test scores in mathematics, reading and science at tenth and twelfth grade. Psychology, Evolution & Gender, 2(2): 119-125.

Moritz, B., Hill, A.V., Donohue, K. (2013) Individual differences in the newsvendor problem: Behavior and cognitive reflection. Journal of Operations Management, 31(1-2): 72-85.

Nagel, R. (1995) Unraveling in guessing games: An experimental study. The American Economic Review, 85(5): 1313-26.

Noussair, C.N., Trautmann, S.T., van de Kuilen G. (2014) Higher order risk attitudes, demographics, and financial decisions. The Review of Economic Studies, 81 (1): 325-355.

Oechssler, J., Roider, A., Schmitz, P.W. (2009) Cognitive abilities and behavioral biases. Journal of Economic Behavior & Organization, 72(1): 147-152.

Paolacci, G., Chandler, J., Ipeirotis, P.G. (2010) Running experiments on Amazon Mechanical Turk. Judgment and Decision Making, 5(5): 411–419.

Pennycook, G., Cheyne, J., Seli, P., Koehler, D., Fugelsang, J. (2012) Analytic cognitive style predicts religious and paranormal belief. Cognition, 123: 335– 346.

Peterson, R.A. (2001) On the use of college students in social science research: Insights from a second-order meta-analysis. Journal of Consumer Research, 28(3): 450-461.

Peysakhovich, A., Rand, D.G. (2015) Habits of virtue: Creating norms of cooperation and defection in the laboratory. Management Science, doi:10.1287/mnsc.2015.2168

Ponti, G., Rodriguez-Lara, I. (2015) Social preferences and cognitive reflection: Evidence from dictator game experiment. Frontiers in Behavioral Neuroscience, doi: 10.3389/fnbeh.2015.00146

Riedel, J.A., Nebeker, D.M., Cooper, B.L. (1988) The influence of monetary incentive on goal choice, goal commitment, and task performance. Organizational Behavior and Human Decision Processes, 42: 155-180.

Scott, W. E., Farh, J.L., Podsakoff, P.M. (1988) The effects of 'intrinsic' and 'extrinsic' reinforcement contingencies on task behavior. Organizational Behavior and Human Decision Processes, 41: 405-425.

Sloman, S.A. (1996) The empirical case for two systems of reasoning. Psychological Bulletin, 119: 3–22.

Stanovich, K.E., West, R.F. (2000) Individual differences in reasoning: Implications for the rationality debate? Behavioral and Brain Sciences, 23: 645-665.

Toplak, M.E., West, R.F., Stanovich, K.E. (2011) The Cognitive Reflection Test as a predictor of performance on heuristics and biases tasks. Memory & Cognition, 39: 1275-1289.

Toplak, M.E., West, R.F., Stanovich, K.E. (2014) Assessing miserly information processing: An expansion of the Cognitive Reflection Test. Thinking & Reasoning, 20(2): 147-168.

Weiss, V. (1986) From memory span to the quantum mechanics of intelligence. Personality and Individual Differences, 7: 737-749.

Appendix

Figure A1. Screenshot of the Cognitive Reflection Test survey

Cognitive Reflection Test survey

It would be greatly appreciated if you could fill out the below survey regarding your own research using the CRT. If you had MULTIPLE studies please fill out a second, third, etc. survey after the first one.

Pablo Branas Garza, Praveen Kujal & Balint Lenkei, Middlesex University.

* Required

Please provide us the name of the authors, title, year and details about the journal (if published) in order for us to properly cite it. *
(If not published please just state "unpublished")

Contact e-mail address *

Location of the study (city and country) *

(1) Total Number of CRT participants *

Among those mentioned in (1) how many were female?
(If you did not register gender please state it)

(2) How many of the total answered the BAT AND BALL question correctly? *

Among those mentioned in (2) how many were female?
(If you did not register gender please state it)

(3) How many of the total answered the MACHINES question correctly? *

Among those mentioned in (3) how many were female?
(If you did not register gender please state it)

(4) How many of the total answered the LILLY PAD question correctly? *

Among those mentioned in (4) how many were female?
(If you did not register gender please state it)

(5) Out of the total how many participants answered all THREE questions correctly?

Among those mentioned in (5) how many were female?
(If you did not register gender please state it)

(6) Out of the total how many participants answered only TWO questions correctly?

Among those mentioned in (6) how many were female?
(If you did not register gender please state it)

(7) Out of the total how many participants answered only ONE question correctly?

Among those mentioned in (7) how many were female?
(If you did not register gender please state it)

Did you pay subjects monetary incentives for correct answers? *

- Yes
- No

Was the CRT an online or a paper and pencil test? *

- Online
- Paper and pencil

What was the order of the CRT questions? *

- Bat and Ball; Machines; Lilly Pad
- Bat and Ball; Lilly Pad; Machines
- Machines; Lilly Pad; Bat and Ball
- Machines; Bat and Ball; Lilly Pad
- Lilly Pad; Bat and Ball; Machines
- Lilly Pad; Machines; Bat and Ball

If you run experiments during the session, was the CRT done before, after or in between the experiments? *

- At the beginning
- Was the last activity
- In between
- Did not run experiments

Any additional information you would like to mention? *
(e.g. experiment was conducted with children)

Submit

Table ER1. Robustness check: Females only

	(1) *B&B*	(2) *Machines*	(3) *Lillypad*	(4) *None*	(5) *1*	(6) *2*	(7) *3*
(1) monetary incentives	0.012	-0.064	0.101*	0.000	-0.016	-0.037	0.053
	(0.057)	(0.064)	(0.053)	(0.066)	(0.021)	(0.030)	(0.034)
(2) student	0.127***	-0.027	0.039	-0.080**	0.031***	0.029	0.020
	(0.036)	(0.026)	(0.041)	(0.039)	(0.009)	(0.018)	(0.022)
(3a) in-between experiments	-0.046	0.009	-0.071	0.047	0.000	-0.015	-0.032
	(0.047)	(0.039)	(0.053)	(0.048)	(0.014)	(0.020)	(0.030)
(3b) after the experiment	-0.045	-0.004	-0.092**	0.064	-0.010	-0.029	-0.025
	(0.043)	(0.035)	(0.042)	(0.043)	(0.010)	(0.018)	(0.026)
(4) visibility	0.009*	0.017***	0.008	-0.007	-0.007***	0.004	0.011**
	(0.005)	(0.005)	(0.005)	(0.005)	(0.002)	(0.002)	(0.004)
(5) standard sequence	0.093**	0.106***	0.151***	-0.149***	0.017	0.059***	0.072**
	(0.041)	(0.033)	(0.044)	(0.044)	(0.013)	(0.018)	(0.028)
(6) computerized	0.069*	0.100**	0.112**	-0.130***	0.032**	0.058***	0.040
	(0.038)	(0.048)	(0.045)	(0.049)	(0.015)	(0.022)	(0.025)
Constant	0.053	0.072	0.039	0.759***	0.228***	0.086**	-0.073*
	(0.069)	(0.073)	(0.068)	(0.080)	(0.027)	(0.036)	(0.044)
N	19995	19995	19995	20945	20945	20945	20945
R^2	0.026	0.020	0.032	0.031	0.005	0.009	0.013

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, *p<0.1. The regressions also controls for the country of the study by using two dummy variables: europe and anglo-saxon.

Table ER2. Robustness check: excluding studies where the experimenters used monetary incentives to reward correct answers

	(1) *B&B*	(2) *Machines*	(3) *Lillypad*	(4) *None*	(5) *1*	(6) *2*	(7) *3*
(1) female	-0.107***	-0.176***	-0.197***	0.181***	0.004	-0.066***	-0.118***
	(0.011)	(0.010)	(0.011)	(0.010)	(0.007)	(0.008)	(0.008)
(2) student	0.108***	-0.037	0.046	-0.066*	0.017**	0.027*	0.022
	(0.036)	(0.024)	(0.040)	(0.035)	(0.007)	(0.014)	(0.023)
(3a) in-between experiments	-0.070	-0.055*	-0.115**	0.083*	0.012	-0.026	-0.069**
	(0.050)	(0.030)	(0.056)	(0.045)	(0.010)	(0.016)	(0.032)
(3b) after the experiment	-0.065	-0.047	-0.123***	0.088**	-0.004	-0.033**	-0.051*
	(0.040)	(0.033)	(0.043)	(0.038)	(0.008)	(0.014)	(0.027)
(4) visibility	0.016***	0.024***	0.011**	-0.011**	-0.009***	0.004	0.016***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.002)	(0.002)	(0.004)
(5) standard sequence	0.119***	0.122***	0.162***	-0.153***	0.006	0.053***	0.094***
	(0.039)	(0.032)	(0.041)	(0.040)	(0.011)	(0.015)	(0.029)
(6) computerized	0.063*	0.112**	0.154***	-0.132***	0.016	0.066***	0.051
	(0.036)	(0.047)	(0.044)	(0.040)	(0.013)	(0.018)	(0.032)
Constant	0.271***	0.390***	0.365***	0.428***	0.246***	0.177***	0.149**
	(0.088)	(0.073)	(0.084)	(0.086)	(0.026)	(0.039)	(0.062)
N	31766	31766	31766	33338	33338	33338	33338
R^2	0.051	0.063	0.077	0.072	0.005	0.016	0.046

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, *p<0.1. The regressions also controls for the country of the study by using two dummy variables: europe and anglo-saxon.

Table ER3. Robustness check: Regressions with non-student samples only

	(1) *B&B*	(2) *Machines*	(3) *Lillypad*	(4) *None*	(5) *1*	(6) *2*	(7) *3*
(1) female	-0.086***	-0.154***	-0.178***	0.165***	-0.007	-0.061***	-0.097***
	(0.010)	(0.013)	(0.014)	(0.012)	(0.008)	(0.010)	(0.008)
(2) monetary incentives	0.073	0.066	0.189**	-0.108	-0.023**	0.058***	0.072
	(0.090)	(0.097)	(0.070)	(0.088)	(0.008)	(0.020)	(0.073)
(3a) in-between experiments	-0.051	-0.044	-0.093*	0.077*	0.017**	-0.026*	-0.069*
	(0.046)	(0.031)	(0.048)	(0.039)	(0.008)	(0.014)	(0.035)
(3b) after the experiment	-0.032	-0.004	-0.049	0.045	-0.003	-0.011	-0.031
	(0.023)	(0.048)	(0.042)	(0.042)	(0.007)	(0.018)	(0.028)
(4) visibility	0.040***	0.028	0.018	-0.022	-0.010***	0.003	0.029***
	(0.006)	(0.017)	(0.015)	(0.014)	(0.002)	(0.006)	(0.009)
(5) standard sequence	0.189***	0.156***	0.199***	-0.203***	-0.004	0.061***	0.146***
	(0.038)	(0.038)	(0.044)	(0.040)	(0.012)	(0.016)	(0.031)
(6) computerized	-0.030	0.012	0.053	-0.057	0.041***	0.035	-0.020
	(0.065)	(0.066)	(0.057)	(0.063)	(0.005)	(0.022)	(0.051)
Constant	-0.193**	0.088	0.006	0.817***	0.278***	0.085**	-0.180**
	(0.073)	(0.111)	(0.095)	(0.097)	(0.016)	(0.037)	(0.069)
N	21983	21983	21983	23199	23199	23199	23199
R^2	0.041	0.044	0.078	0.071	0.007	0.017	0.042

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, *p<0.1. The regressions also controls for the country of the study by using two dummy variables: europe and anglo-saxon.

Table ER4. Robustness check: excluding the studies where the researchers did not run experiments

	(1) *B&B*	(2) *Machines*	(3) *Lillypad*	(4) *None*	(5) *1*	(6) *2*	(7) *3*
(1) female	-0.107***	-0.167***	-0.186***	0.170***	0.009	-0.063***	-0.116***
	(0.012)	(0.008)	(0.010)	(0.010)	(0.008)	(0.007)	(0.009)
(2) monetary incentives	0.046	0.091*	0.110*	-0.081	-0.007	0.019	0.069*
	(0.045)	(0.051)	(0.059)	(0.052)	(0.017)	(0.021)	(0.039)
(3) student	0.108***	-0.026	0.051	-0.070	0.014	0.030*	0.026
	(0.037)	(0.029)	(0.049)	(0.042)	(0.009)	(0.017)	(0.024)
(4a) in-between experiments	-0.130**	-0.039	-0.140**	0.101**	0.015	-0.022	-0.093**
	(0.055)	(0.048)	(0.058)	(0.050)	(0.015)	(0.020)	(0.036)
(4b) after the experiment	-0.109***	-0.037	-0.135***	0.095**	0.005	-0.029	-0.071**
	(0.041)	(0.048)	(0.047)	(0.041)	(0.012)	(0.019)	(0.028)
(5) visibility	0.002	0.013*	0.005	-0.004	-0.004	0.004	0.004
	(0.007)	(0.007)	(0.009)	(0.008)	(0.003)	(0.003)	(0.006)
(6) standard sequence	0.120***	0.115***	0.175***	-0.164***	0.018	0.057***	0.089***
	(0.038)	(0.031)	(0.041)	(0.038)	(0.012)	(0.015)	(0.029)
(7) computerized	0.084**	0.136***	0.145***	-0.130***	0.000	0.055***	0.076***
	(0.035)	(0.038)	(0.048)	(0.041)	(0.010)	(0.021)	(0.024)
Constant	0.392***	0.441***	0.396***	0.399***	0.195***	0.170***	0.237***
	(0.117)	(0.100)	(0.119)	(0.114)	(0.033)	(0.047)	(0.082)
N	28268	28268	28268	28624	28624	28624	28624
R^2	0.056	0.068	0.086	0.086	0.002	0.019	0.048

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, *p<0.1. The regressions also controls for the country of the study by using two dummy variables: europe and anglo-saxon.

Table ER5. Robustness check: excluding those studies where the experimenters used Amazon Mechanical Turk for the tests

	(1) *Bat and Ball*	(2) *Machines*	(3) *Lillypad*	(4) *None*	(5) *1*	(6) *2*	(7) *3*
(1) female	-0.115***	-0.181***	-0.202***	0.180***	0.014**	-0.070***	-0.124***
	(0.012)	(0.011)	(0.011)	(0.011)	(0.007)	(0.008)	(0.009)
(2) monetary incentives	-0.022	0.006	0.045	-0.010	-0.002	0.002	0.010
	(0.043)	(0.042)	(0.047)	(0.043)	(0.014)	(0.017)	(0.036)
(3) student	0.171***	0.033	0.095**	-0.113***	0.001	0.031**	0.081***
	(0.041)	(0.032)	(0.046)	(0.042)	(0.009)	(0.015)	(0.029)
(4a) in-between experiments	-0.033	0.019	-0.088*	0.054	-0.010	-0.020	-0.023
	(0.047)	(0.041)	(0.049)	(0.042)	(0.015)	(0.016)	(0.035)
(4b) after the experiment	-0.030	-0.001	-0.093*	0.055	-0.008	-0.025	-0.022
	(0.045)	(0.035)	(0.047)	(0.043)	(0.009)	(0.015)	(0.032)
(5) visibility	0.003	0.010*	0.001	-0.002	-0.005**	0.002	0.005
	(0.006)	(0.006)	(0.006)	(0.005)	(0.002)	(0.002)	(0.005)
(6) standard sequence	0.059	0.059	0.118**	-0.121***	0.042**	0.042**	0.038
	(0.041)	(0.036)	(0.046)	(0.045)	(0.016)	(0.019)	(0.029)
(7) computerized	0.032	0.084*	0.106**	-0.095**	0.014	0.049**	0.032
	(0.040)	(0.049)	(0.052)	(0.046)	(0.011)	(0.020)	(0.034)
Constant	0.248***	0.339***	0.333***	0.499***	0.200***	0.167***	0.134**
	(0.077)	(0.075)	(0.079)	(0.081)	(0.025)	(0.035)	(0.054)
N	31200	31200	31200	31870	31870	31870	31870
R^2	0.049	0.057	0.068	0.064	0.003	0.013	0.043

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, *p<0.1. The regressions also controls for the country of the study by using two dummy variables: europe and anglo-saxon.

Table ER6. Robustness check: Excluding studies where the sequence of questions was randomized

	(1) *B&B*	(2) *Machines*	(3) *Lillypad*	(4) *None*	(5) *1*	(6) *2*	(7) *3*
(1) female	-0.117***	-0.176***	-0.196***	0.176***	0.012**	-0.065***	-0.124***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.034)	(0.000)	(0.000)
(2) computerized	0.021	0.077	0.098*	-0.087*	0.013	0.048**	0.026
	(0.601)	(0.118)	(0.062)	(0.060)	(0.274)	(0.017)	(0.443)
(3) student	0.111***	-0.017	0.046	-0.063*	0.011	0.022*	0.030
	(0.003)	(0.520)	(0.272)	(0.079)	(0.183)	(0.083)	(0.251)
(4a) in-between experiments	-0.055	0.000	-0.097*	0.064	-0.003	-0.019	-0.043
	(0.266)	(0.995)	(0.073)	(0.147)	(0.845)	(0.231)	(0.242)
(4b) after the experiment	0.007	0.027	-0.050	0.017	-0.010	-0.013	0.006
	(0.859)	(0.371)	(0.193)	(0.606)	(0.236)	(0.305)	(0.826)
(5) standard sequence	-0.031	-0.087*	-0.044	0.024	0.047*	0.001	-0.072**
	(0.524)	(0.068)	(0.288)	(0.664)	(0.092)	(0.956)	(0.016)
(6) monetary incentives	-0.005	0.019	0.060	-0.025	-0.003	0.005	0.023
	(0.918)	(0.690)	(0.224)	(0.584)	(0.867)	(0.772)	(0.570)
(7) visibility	0.007	0.014**	0.004	-0.004	-0.007***	0.002	0.009*
	(0.214)	(0.023)	(0.532)	(0.424)	(0.004)	(0.281)	(0.096)
Constant	0.317***	0.460***	0.475***	0.373***	0.204***	0.201***	0.222***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	32846	32846	32846	34418	34418	34418	34418
R^2	0.037	0.049	0.053	0.048	0.003	0.01	0.036

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, *p<0.1. The regressions also controls for the country of the study by using two dummy variables: europe and anglo-saxon.

Conclusions

The thesis summarises the main findings of my doctoral studies at Middlesex University titled 'Essays on Human Capital'. It follows a top down approach by first discussing macroeconomic issues related to human capital and later looking into two specific micro level applications of human capital including mother's educational expectations in developing countries and drivers of performance on a cognitive ability test.

Human capital present in the labour force enhances labour productivity which leads to a higher equilibrium level of output. In turn, human capital has several determinants including education, health, cognitive and non-cognitive abilities. The first chapter investigates one of the channels through which human capital may affect macroeconomic growth by looking at both the direct impact of the investments in education and the way educational resources are distributed within the society. The study recognises that it is not just the accumulation of human capital to play a role in determining growth but also the way human capital is distributed.

The first part of the analysis relies on traditional growth models contrasting the conventional average year of schooling measures (BL,CS and PWT) with PISA test scores and on a measure of educational inequality based on the Gini coefficient. The second part looks at the long run relationship of human capital accumulation and educational inequality on growth using a method that encompasses previous specifications and estimates both short-run and long run effects.

Overall, the results provide evidence that over the last few decades human capital, measured by education, has had an impact on economic growth. This is particularly true when using dynamic specification and controlling for cross-sectional heterogeneity and unknown common factors. The findings also support the view that governmental policies should focus on allocating national resources to education improving both accessibility to basic education in all parts of the society while at the same time improving the quality of education systems.

Although resources spent on education are important to boost growth in developing countries, there are other issues derived from individuals' preferences and behaviour that prevent efficient investment in education. One of these channels is parental expectations and their beliefs about their children. Parental expectations have substantial economic and social impact on the household. High expectations of parents lead children to set higher standards for their education and to make greater demands on themselves from an early age which in turn results in high achievement, better attendance and more positive attitudes towards school.

In particular, the second chapter looks at how parents' time discounting and degree of risk aversion impact expectation about their children's educational attainment in a developing country setting. Education in developing countries is hypothesised to be a risky investment for households, while returns are only realised after some time in the future. Data from a field experiment in Peru was used to analyse the relationship between parental expectations on children education attainment and parental risk and time preferences among rural households.

After controlling for a set of household, parental and child characteristics it is found that the risk-aversion parameter (sigma) is significantly and positively related to expectations on schooling choices. This suggest the mothers in this study do not consider education as a risky investment in rural Peru. Instead, they perceive education as a safe option and fear of losing out on higher returns resulting from high educational attainment.

It is also found that impatient parents (higher discount rate) tend to have lower expectations compared to patient parents. This is due to parents with high discount rate are expecting immediate financial gains from educational attainment of their children, and therefore they may tend to underinvest given that they are not patient enough to wait for the return. These results overall support the view that parental preferences may play a role in preventing efficient investment in childrens education in developing countries.

As mentioned earlier, human capital has several aspects including, among others, cognitive abilities. The channels through which cognitive abilities promotes economic growth is mediated through multiple channels, for example, greater technological competitiveness, lower fertility, reduced burden of infectious diseases and increased domestic savings rates, whereas at the micro level cognitive abilities are said to explain individual decisions and behaviour.

However, there is a lack of consensus in the literature on the best way of measuring it. Education has been often used as a proxy for cognitive abilities given that these abilities are not easily observed. There are an increasing number of alternative measures proposed that are gaining popularity, one of which is the Cognitive Reflection Test (CRT).

The CRT was first proposed by Frederick (2005) and since then has been extensively used in the experimental economics and psychology literature. Frederick proposed the test based on a dual-system theory made up of two cognitive processes: System 1, executed quickly without much reflection and System 2, more deliberate and requiring conscious thought and effort. The questions in the CRT have an immediate (intuitive) incorrect response (System 1). However, the correct response requires some deliberation, i.e. the activation of System 2.

The CRT's use as a covariate to explain behavior and decisions has significantly increased in the past few years and experiments have shown its usefulness in predicting behavior. However, little is known about if the test is gender biased, whether incentives matter or how different implementation procedures impact outcomes. Therefore, in chapter three I present the results of a meta-study of 118 CRT studies comprising of 44,558 participants across 21 countries.

The results show that there is a negative correlation between being female and the overall, and individual, correct answers to CRT questions, whereas monetary incentives do not impact performance. Regarding implementation procedures, taking the test at the end of the experiment negatively impacts performance, while students generally perform better compared to non-students. The results also suggest mixed evidence on whether the sequence of questions matters or whether computerised tests improve results. Overall, the findings imply that the implementation of the test and the subject pool may bias test results, which in turn can have an impact on behavioral or macro growth models.