Returns to Education and Skills in the non-OECDs

Evidence from Urban Kenya

A thesis submitted
for the degree of Doctor of Philosophy in Economics
at the University of Glasgow

by

Nnanna Osita Oledibe

September 2023

Abstract

This thesis examines the returns to education and skills of the labour force (aged, 15-64) in urban Kenya, a non-OECD[1] country. Comprised of four self-contained analytical chapters with interconnected themes, this thesis uses the World Bank's STEP[2] Household Survey for Kenya. Beyond reduced-form equation modelling, this study deploys a structural equation modelling approach. Hence, raising the internal and external validity of return estimates. Using the 1985 curriculum structural reforms in Kenya, this study exploits exogenous variations in schooling and skill from which causal inference is drawn, providing empirical evidence that informs policymaking. In this study, the consideration for skill as opposed to mere schooling[3] presents a novel approach to examining human capital, particularly, this unravels useful insights that improve existing understanding of the mechanisms through which schooling raises skill, and in turn, other labour market outcomes such as earnings.

For the first analytical chapter — Chapter 2: *Educational Attainment and Skill Proliferation* — In examining the effects of schooling on skills in Kenya[4], the evidence suggests substantial 'inefficiency in schooling' in urban Kenya. The term 'inefficiency in schooling' describes a state where workers with relatively low educational attainment have relatively high reading proficiency. To give more understanding to this phenomenon, I stress the importance of the effects of background characteristics[5] on access to schooling (and skill). Findings provide a basis for an argument for more effort to incentivise equity in access to schooling for all, over recent arguments for increased quality in schooling. I argue that whilst the latter is of 'noble aim', resource constraints in education provision mean efforts towards increased quality over equity-in-access-to-schooling inhibit skill proliferation, particularly in developing contexts. Exploiting the 1985 curriculum reform in a Difference-in-

[1] The OECDs (Organisation for Economic Co-operation and Development) is a group of thirty-eight (38) **high-income** countries including the United States of America and the United Kingdom. Please, see the full list here https://www.oecd.org/about/. On the other hand, the non-OECDs here mean **low- and mid-income countries** of which the sub-Sahara Africa is part of.

[2] STEP is an abbreviation for Skills Toward Employment and Productivity.

[3] Schooling is taken to mean 'time' spent in formal education.

[4] Kenya is part of the sub-Sahara Africa known to have the least schooling and skills, relative to the other regions of the world.

[5] Background characteristics are proxied with parental education and socioeconomic status at age 15. The 'advantaged' as used in this study, are those (respondents) that have a father with post-secondary schooling, in most cases, these have high socioeconomic status at age 15. On the other hand, the 'disadvantaged' are respondents who have a father without post-secondary schooling, in most cases these have low socioeconomic status at age 15.

Differences (DiD) analysis, the evidence suggests an upward mobility in schooling for the ‘disadvantaged’ (respondents that have fathers without post-secondary schooling). However, no useful evidence (from which causal inference can be drawn) on skill is attributable to parental post-secondary education or socioeconomic status at age 15. Further evidence suggests this inconsistency in the effects of background characteristics on schooling and skill is partly due to the DiD estimator that gives the Average Treatment Effects on the Treated (ATET); and the inefficiency in education, in urban Kenya. Interestingly, turning to the Two-Stage Least Squares Instrumental Variables (2SLS-IV approach which gives the Local Average Treatment Effects (LATE), the evidence suggests that relative to the ‘advantaged’ (those that have fathers with post-secondary education), for the ‘disadvantaged’, the effect of an additional year of schooling on skill is positive and statistically significant. Although quite different, in that the LATE in this case captures the effect of schooling on skill strictly for those impacted by the reform; on the other hand, the ATET, captures the effects of parental education on schooling and skill, regardless of the reform. I find outcomes of the ATET and LATE complementary and strongly responsive to the 1985 curriculum reform and the prevailing inefficiency in schooling in urban Kenya. Particularly, having the coefficients (ATET and LATE) indicating a substantial rise in the schooling and skills of the ‘disadvantaged’ is strongly attributable to the reform and the inefficiency in schooling. This is not to suggest the ‘disadvantaged’ have higher schooling and skills relative to the ‘advantaged’ as the reverse is the case, with evidence of substantial skill differential attributable to differences in schooling endowments between the ‘advantaged’ and the ‘disadvantaged’. However, as earlier highlighted, the evidence suggests, relative to the ‘advantaged’ that the reform (and the inefficiency in schooling) drives the schooling and skill of the ‘disadvantaged’. Ultimately, whilst the evidence suggests the effects of inefficiency in schooling looms large and should be addressed speedily possibly by addressing the quality needs, particularly, at higher levels of schooling (above ISCED2[6]), further evidence from the 2SLS-IV approach suggests, a more positive and substantial effect of an additional year of schooling on the skill, for all impacted by the reform (regardless of background characteristics). This suggests efforts aimed at raising equity in access to schooling should not be discouraged out of quality concerns. Hence, I argue that reforms that incentivise access to schooling for increased educational attainment are more crucial for

[66] Please, see the data subsection of the first analytical chapter for credential categories. The ISCED2 credential category represents the credential category of the employed who attained lower-secondary education. This is equivalent to an average of eight (8) years of schooling, in urban Kenya.

skill proliferation than efforts to raise the quality of school inputs. Besides, raising the quality of school inputs can inhibit access to schooling, due to resource constraints.

In the second analytical chapter — Chapter 3: *Private Returns to Education and Skills* — I estimate the private returns to education and skills and examine the wage differential across gender and employment categories. The findings suggest, controlling for schooling, the OLS (Ordinary Least Squares) return estimates of non-cognitive skills are robust, with Openness to Experience and Conscientiousness yielding positive and statistically significant wage effects from which causal inference is drawn. Openness to Experience has the strongest effect with a standard rise in Openness explaining a 35.9% rise in hourly earnings, statistically significant at the 0.1% level. However, a standard rise in Conscientiousness explains a 12.6% rise in hourly earnings, statistically significant at the 5% level. The 2SLS-IV estimates show consistent estimates of returns to schooling and cognitive skills. Findings suggest, no evidence of statistically significant wage effects of schooling and skill from which causal inferences are drawn. Further evidence from subsampling (heterogeneity analysis) shows that relative to the female gender, the male gender has positive returns to their schooling and cognitive skills, with an additional year of schooling explaining a 25.6% rise in hourly wage, statistically significant at the 1% level. For the measure of cognitive skills (reading proficiency, unstandardised), the evidence suggests that a unit rise in reading proficiency in PV (Plausible Values) explains a 0.77% rise in hourly wage, statistically significant at the 5% level. Using the first stage of the Oaxaca-Blinder decomposition as the baseline estimates, the evidence suggests a 23% hourly wage differential across genders in urban Kenya. Differences in schooling and skills characteristics/endowments explain about 37% of the wage difference across genders. Further evidence suggests that a substantial proportion of the wage differential between genders is due to (potential) discrimination. Particularly, females are not discriminated against based on their cognitive skills or schooling but rather, the evidence suggests the potential discrimination in wage between genders comes through differences in their non-cognitive skills (or personality traits), specifically, whilst the males are better rewarded for their Openness to Experience; the females are better rewarded for their hard work (Conscientiousness). Substantial policy insights abound in these outcomes.

In the third analytical chapter — Chapter 4: *Human Capital Externalities and Social Returns* — I examine pecuniary and non-pecuniary human capital externalities[7]. Evidence from OLS output suggests substantial negative externalities of schooling in urban Kenya. Specifically, the findings show negative pecuniary and non-pecuniary externalities (of schooling) that become less negative with rising aggregate (district-level) schooling. Hence, the negative pecuniary and non-pecuniary externalities of schooling become non-negative (positive) at a certain level of aggregate schooling. This is consistent with the argument for more schooling (over quality inputs), as in Chapter 2—the first analytical Chapter. Interestingly, findings show the pecuniary externalities of skills are positive and statistically significant. The differences in the externalities of schooling and skill unravel interesting insights that question some 'stylised' facts in the literature. Particularly, findings strongly suggest aggregate schooling is meaningful (or makes economic sense) only at a certain threshold, and any level below this threshold inhibits earnings and skill proliferation. Interestingly, on the other hand, increasing aggregate skill (regardless of the skill level) has favourable effects on earnings.

Finally, in the fourth analytical chapter — Chapter 5: *Returns to Education and Skill in a Dynamic Framework*. Amidst limitations that accrue from using a cross-section of data, the main objective of this chapter is to test the robustness of estimates from previous chapters (chapters 2, 3, and 4) that use a single cross-section of data. Hence, a dynamic framework that accounts for data limitations and supports causal identification helps to raise the external and internal validity of estimates addressing possible biases in return estimates. Inspired by the study of Krishnakumar and Nogales (2020) in Bolivia, I deploy the Technology of Skills formation—a dynamic framework, pioneered by Cunha and Heckman (2007). Overall, findings from the Structural Equation Models (SEMs) are consistent with estimates in reduced form, as in the previous chapters. The SEMs further unravel some useful insights that improve understanding of the return estimates. Particularly, the findings from the SEMs improve understanding of the Difference-in-Differences analysis of Chapter 2 that show father's post-secondary education impacts the schooling but not the skills of the offspring. Deploying the dynamic framework (SEMs) affirms having a father with post-secondary education not only explains the schooling, but the skills of their wards. These findings present substantial evidence of at least persistence (and upward mobility) in skill and education between parents and their offspring. This strongly accentuates an intergenerational

[7] The former, pecuniary externalities, entail an examination of aggregate schooling and skill (across districts) on individual wage. The latter, non-pecuniary externalities involve an examination of aggregate schooling (across districts) on individual skill.

transmission mechanism for educational attainment and skill proliferation that should not be overlooked in education, skills, and employment policymaking. This finding in sub-Saharan Africa is consistent with related findings from studies in the OECDs.

Table of Contents

List of Tables

List of Figures

List of Appendices

Acknowledgement

To God be the glory—for in Him I live, move, and have my being. I thank God Almighty for His great and wonderful deeds.

This thesis is a product of years of research—thinking, writing, extensive reviews of the literature, and data analytics using STATA and R. I am grateful to my advisors and managers: Professor Kristinn Hermannsson, Professor Georgios Panos, Professor Jeanette Findlay, Professor Georgios Sermpinis and Professor Evangelos Vagenas-Nanos, for their unwavering academic support and fuel that kept me up and on track. I look forward to a long and fruitful relationship with them. This thesis benefitted immensely from the support of several academics of the College of Social Sciences. I am indebted to Dr Ioana Moldovan who was there all through the years, together with her team of Economics PGR conveners, Dr Arjunan Subramanian, and Dr Francesca Flamini. I am grateful to members of the Applied Economic Research Cluster who reviewed my work and gave useful feedback over the years. I am immensely grateful to Dr Anwen Zhang for his reviews, research advisory, and for co-organising (with Prof. Eugenio Proto) interesting seminars. I thank Dr Tanya Wilson for her reviews and feedback. I am indebted to Professor Michele Battisti for his reviews and feedback, and for being a research luminary. I am very grateful to Dr Theodore Koutmeridis for his great help at the (turbulent) commencement of my research. I thank Dr Rhys Crilley of the School of Social & Political Sciences who offered tremendous support and mentoring, meeting up at the JMSLH (James McCune Smith Learning Hub) at 9:30 am fortnightly during the writing up phase of this work, in Summer 2022.

I want to thank our ever-ready and very helpful research administrators, Ms. Sophie Watson, Mrs Christine Athorne, and Mrs Angela Foster for their great support – completing all paperwork for annual reviews, registration, funding applications, travels, supervisory sessions, and many more administrative functions – that kept me running through the years. Special thanks go to our motherly Christine who retired last year.

Learning with peers at this level cannot be over-emphasised. Having an active Reading Group is a great support for us for which I am grateful. I am seizing this opportunity to thank my great colleagues in the Economics PhD programme. To mention a few, I would like to thank Tren Ma who was there for me from my first year and all through my studentship—shared the same office space in my final year. I also owe gratitude to Max Schroders and

Prateek Bhan for being supportive in my early years, not forgetting Michael Vallely, a great mind who was there with me in my first year.

Finally, my supportive family. This is to thank my mother Mrs Beatrice Oledibe, a schoolteacher who later became a school administrator. She took on the challenge of actualising my late father's aspiration of providing a good education to nine (9) children in one of the most deprived towns in Lagos, Nigeria. Her dedication and selflessness are second to none. I thank my lovely siblings for their gifts, words of encouragement, and prayers. Specifically, I would like to thank my sister, Ms Chinonso Oledibe who is referred to as my mother, by my mother. I say thank you for those nights and early mornings together. I dedicate this thesis to my late father — Mr Michael Oledibe Sr, a schoolteacher who later ran a chemist — who was a gift to me in my key formative years. He was there for only the first thirteen years of my life but prepared me for life! He made me understand the dignity of every honest and hard work.

Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, this dissertation is the result of my work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Name: Nnanna Osita Oledibe

Signature:................................

1 Introduction to the Thesis

Besides the non-pecuniary effects of education e.g., reduction in crimes (Machin et al., 2011) and increased health outcomes (Campbell et al., 2014), there is a burgeoning literature on the wage effects of investment in education. It is understood that education impacts wages through its impact on skills known to raise productivity. Governments, Development Finance Institutions, and other Third-Sector organisations use education as a tool for several growth agendas, these include, poverty alleviation, equality, and new technology, to mention a few. However, central to the ideas of education-for-growth, is the objective to achieve useful labour force participation, which makes employability (skill proliferation) a crucial mechanism through which education yields useful outcomes for economic growth (Nelson and Phelps, 1966) and development. As a strategy for growth and development, countries use education policy (or related reforms) to influence human capital for gainful employment. Hence, education, skill, and labour market outcomes such as earnings, are known to have strong links (Krishnakumar and Nogales, 2020) that can explain economic growth and development. Whilst a useful understanding of this link (schooling-skill-earnings) exists in the OECDs[8], relatively, less attention has been given to examining the link in the non-OECDs. This study unravels this link in a non-OECD context. Hence, the primary (or overarching) objective of this study is to respond to the question:

What is the link (or clear relationships) across education-skill-labour market outcomes in the non-OECDs?

With a focus on Kenya[9], this study exploits the World Bank's Household STEP data for Kenya, improving understanding of some fundamental relationships in the Economics of Education and Labour Economics particularly as they relate to economic growth and

[8] Owing to the plethora of extensive studies in microeconomics, theory, and macroeconomics, in the OECDs (see Sianesi and Reneen, 2003; Carneiro et al 2011; and Harmon et al., 2003).

[9] There are several reasons governments (of countries) invest in education. The expectation that education improves the wealth, well-being, and welfare of the educated and the economy in aggregate is a major rationale for government intervention. Insights that inform the optimal levels of government involvement (especially through policies and reforms) in education will result in useful economic returns to investments in education. Without government intervention in education in the non-OECDs, education in most non-OECDs will be a privilege to the few who can afford it, this peculiarity of most non-OECDs necessitates government intervention in the region. In this study, the choice of Kenya (a country in sub-Saharan Africa) is influenced by several factors. Over the years, as strategies for economic growth and development, the government of Kenya is known to have taken useful steps in developing employment (particularly, self-employment) and education reforms. Hence, the case of Kenya provides a useful testbed in examining the schooling-skill-earnings link as earlier raised.

development in the non-OECDs. To address biases, and hence raise the internal validity of estimates across the analytical chapters of this work, this study exploits a natural experiment—using exogenous variation in schooling and skill from the 1985 curriculum reform, in Kenya—that supports the quasi-experimental approaches deployed to draw causal inferences. It is typical for studies on returns to education to consider either measures of schooling or skill, but it is seldom both. This study has explored both schooling and skill as measures of human capital, examining arguments (see Pritchett, 2001) that suggest 'schooling does not necessarily mean skilling', in most developing contexts, particularly, in sub-Sahara Africa. This study adds to ongoing conversations, providing new insights that relate to sub-Sahara Africa, contributing new empirical evidence to the literature, particularly, improving understanding of the mechanisms through which education results in economic growth by raising skills and other labour market outcomes such as earnings. Beyond contributions to the literature, this study also draws crucial policy conclusions that relate to education, and employment in developing contexts.

The 'overarching question' is fully analysed across the four analytical chapters as earlier discussed[10]. This thesis encompasses the seminal work of Pritchett (2001) 'Where Has All the Education Gone?' Lant Pritchett presents three propositions central to key arguments in the Economics of Education; and Labour Economics, in the non-OECDs.

Pritchett (2001) puts forward the following propositions:

1. The newly created educational capital has gone into piracy; that is, privately remunerative but results in socially unproductive activities – Are Cognitive Skills Applied to Socially Productive Activities?

2. There has been slow growth in the demand for educated labour, so the supply of educational capital has outstripped demand and returns to schooling have declined rapidly – Stagnant Demand for Educated Labour.

[10] (As in the Abstract). I do this by first examining the relationship between education and skills on one hand; and on the other hand, examining the impacts of education/skills on earnings. Next, I consider externalities by examining the effects of aggregate schooling and skill on individual earnings; and the effects of aggregate schooling on individual skill. To examine these links simultaneously, I turn to a structural equation modelling (Technology of Skill Formation) approach. Doing this further raises the external validity of estimates and test the robustness of findings in the first two analytical chapters that exploit reduce-form equation modelling approach. Hence, this study explores two key frameworks—the human capital framework in reduced form models; and a more dynamic framework with structural equation modelling.

3. The education system has failed, so a year of schooling provides few (or no) skills – Did Schooling Create Skills?

The testable predictions of the analytical chapters (chapters 2-5) are carefully analysed and discussed to demystify the links across schooling, skills, and labour market outcomes in sub-Saharan Africa. This entails a keen investigation (in the context of Kenya) of each of the propositions of Pritchett (2001), albeit, in no order.

A useful start to exploring the schooling-skill-earnings link is to examine the relationship between the schooling and skill of the labour force. The schooling-skill relationship is discussed as the 'private non-pecuniary returns to schooling' in Chapter 2, this provides a useful basis for Chapter 3 where the wage effect of schooling and skill is discussed. Chapter 4 examines growth effects by exploring the non-pecuniary externalities of education and the pecuniary externalities of education and skill. Here (in Chapter 4), education and skill are examined at the aggregate or district level. This gives insights into Social Returns to human capital, which is not only an indicator of growth but deemed a useful basis for the government's involvement (hence, the government's investment) in schooling as a public good. Finally, in Chapter 5, acknowledging limitations of the use of a single cross-section of data in reduced-form models, to improve on the internal and external validity of estimates, I turn to a dynamic framework to assess the robustness of most estimates in Chapters 2 - 4. To achieve this, I deploy the Structural Equation Modelling—Technology of Skills Formation—accounting for data limitations and providing a useful basis for causal identification of estimates of returns to schooling and skill.

All four analytical chapters in this study are carefully located in the literature. A literature review shows that in recent times measures of schooling (such as years of schooling) and skill (such as reading proficiency, numeracy, and problem-solving) present the two main measures of human capital in economic research. Earlier studies have considered several other measures such as school enrolment and IQ, for schooling and skill respectively. In recent times, whilst a strand of the literature advocates years of schooling[11] (see Harmon and Walker, 2001) as a measure of human capital, for another, the emphasis[12] is on skill (see

[11] Here, schooling is taken to mean, the time (years) spent in formal education.

[12] In recent times, the emphasis on measures of skills in lieu of other measures of educational attainment (such as the number of years of schooling spent in formal education; and mere credentials) is gaining more grounds as measures of skills are deemed more plausible measure of human capital (see Hanushek and Woessmann, 2008) that relates to economic growth, particularly, in developing contexts where schooling is argued to give little or no skill (see Pritchett, 2001).

Hanushek et al., 2013). Furthermore, in support of the human capital theory (see Becker (1962); and Shultz (1963)), findings from most studies in developed contexts suggest skill is a mechanism through which schooling impacts labour market outcomes (see Chevalier et al., 2004). However, limited empirical evidence that tests the human capital theory exists for the non-OECDs. Examining the assumptions of human capital theory in developing contexts is important, with several studies that assert schooling gives little or no skill in developing contexts, particularly, in sub-Saharan Africa (see Pritchett, 2001). As this may suggest, either a case where skill is less of the mechanism through which schooling impacts labour market outcomes (in this case, the human capital theory is not upheld but rather, the signalling theory is); or a case where little or no skill from schooling (see Pritchett 2001) explain the limited[13] labour market outcomes for the employed. The study of Chevalier et al., (2004) like many other studies in developed contexts did not disregard the effects of signalling[14] (see Spence (1973; and 1979) in discussing evidence of the human capital theory (how skills or productivity mediates the effects of schooling on labour market outcomes). Interestingly, evidence from Chevalier et al., (2004) suggests the difference between the wage effects of schooling; and the effects of schooling on skill gives insights into the proportion of the wage effect of schooling that is ‘signalling’. Hence, this suggests both elements of signalling and productivity (skill) in the wage effects of schooling. This study has examined the wage effects of schooling (Chapter 3) and the rate or extent to which schooling explains skill (Chapter 2) providing evidence in developing contexts. Besides the core testable predictions of the individual chapters, examining the interconnected themes of the chapters of this thesis not only tests the propositions[15] of Pritchett (2001) as earlier highlighted but provides similar evidence in sub-Sahara Africa, as the study of Chevalier et al., (2004), in a developed context. Particularly examining the interconnecting themes of the chapters makes it possible to draw useful inferences on signalling.

A further review of frontiers of the literature on the effects of schooling on skill not only suggests schooling does not necessarily translate to skill, particularly in most developing contexts (see Pritchett, 2001) but as a rationale for this, most related studies cite issues of quality in schooling in developing contexts. In a further review of related literature, some

[13] Here, the human capital theory is upheld but a problem of ‘efficiency in schooling’ may potentially inhibit productivity and growth.

[14] Where education is deemed to impact wages by its signalling as opposed to its productivity-enhancing (skill) effects.

[15] Please, see the propositions of Pritchett (2001) below on pages 2/3.

notable empirical works on the effects of increased schooling on economic growth[16] suggest, that there is no consensus on the impacts of education on skills (or productivity) across countries. Using GDP as a growth measure and enrolments as a measure of aggregate schooling, findings from the study of Barro (1991) suggest developing countries can achieve 'convergence' or increased growth levels with the developed countries, with increased human capital through schooling. However, the works of Benhabib and Spiegel (1994) found no growth with increased enrolments but concluded that increased human capital is required for growth. A few years later, Sala-i-Martin (1997) presented a more objective finding, showing the effects of economic growth for enrolment in primary education and none in secondary education. However, in the same year, the works of Pritchett (1997) showed evidence of 'no convergence' of developing with the developed countries. Pritchett's (2001) propositions suggest a 'divergence' due to differences in skills from schooling between the developed and the developing countries. So far, the findings discussed are consistent with the argument of Temple (1999) who finds the impact of education on growth has not been the same across countries. Hence, as opposed to cross-country studies that examine the education-skill-employment-growth link across several or multiple countries, this necessitates country-specific studies that examine the education-skills-employment-growth links in specific contexts. In empirical works of this nature, such region- or country-specific studies are required for more robust outcomes, particularly, for findings to have specific policy implications. The clear argument is that the reliability of evidence from cross-country studies cannot be compared to single-country studies as the outcome from the former (cross-country studies) may not apply to any constituent jurisdictions, ceteris paribus. As an extension to this, methods/variables that aid robust conclusions in an OECD setting may result in inconsistent findings/conclusions in non-OECD settings. These ideas motivate this single-country study. In this study, acknowledging the nature of the developing context necessitates analyses involving measures of schooling and skill for human capital, instead of mere 'schooling'. Furthermore, the nature of the relationships (schooling-skill-earnings-growth) which involve examining human capital in a developing context requires going beyond typical reduced-form models based on the human capital framework. Hence, to test the robustness of estimates in reduced form, this study deploys the Technology of Skills Formation—a dynamic framework—that also helps to mitigate the defects of the limitations

[16] The relationship between schooling and economic growth may give insights on the effects of skill from schooling as education is deemed to impact growth through skills or productivity-enhancement that can only come through skills.

in data which is peculiar to studies in developing contexts; and supports causal identification, hence the dynamic framework raises the external and internal validity of estimates.

In summary, the chapters of this thesis are briefly outlined as thus: Chapter 1 (this chapter) is an introduction to the thesis; next are Chapters 2—5, the Analytical Chapters where the main testable predictions are examined and discussed; finally, the concluding chapter of this thesis, Chapter 6. This outlines the summary of findings and concluding remarks across the analytical chapters, ultimately, responding to the overarching (primary) question (the schooling-skill-earnings link) by discussing the interconnected themes. Finally, it highlights the limitations of the studies, stating implications for policy and future research work.

2 Educational Attainment and Skill Proliferation

2.1 Introduction

The first analytical chapter of this study entails an analysis and discussion of issues of education, educational attainment, and background characteristics as drivers of 'cognitive skills' proxied by 'reading proficiency' in urban Kenya.

2.1.1 The Problem and Objective of the Study

In recent times (1950 – 2010), there has been a rise in schooling (time spent in formal education) across world regions. However, relative to the OECDs[17], the non-OECDs, particularly sub-Sahara Africa remain the region with the least[18] schooling measured by the number of years of schooling. As of 2010, the average schooling in Eastern Europe was 11.84 years, with a mere 0.5% of its adult population without formal schooling. However, in sub-Sahara Africa the average schooling was only 5.5 years, with 32% of the adult population having no formal schooling (Lee and Lee, 2016). Whilst the current trend in schooling and skill[19] across the world suggests evidence of a relatively weak human capital in developing countries, this evidence shows a clear and positive relationship between the level of formal schooling attained and the skill[20] acquired, regardless of the region of the world. Hence, in this study, I argue that regardless of the quality of schooling, the average skill level of a region is largely attributable to the average schooling in the region. However, this argument is somewhat inconsistent with the claims or propositions of most mainstream studies[21] in the literature that suggest years of schooling give little or no skills in the developing world. These mainstream studies attribute their claims to the 'poor quality'

[17] OECDs, Organisation for Economic Co-operation, and Development constitute a group of thirty-eight (38) high-income countries including. Please, see the full list here https://www.oecd.org/about/. Hence, this makes non-OECDs, as used in this study, to mean, low- and mid-income countries, which includes sub-Saharan Africa, which Kenya is a part of.

[18] Please see, Lee and Lee (2016); Barro and Lee, (2013); Cohen and Soto, (2007).

[19] Comparing outcomes of tests of reading proficiency for similar household surveys using the PIAAC (Programme for the International Assessment of Adult Competencies) for the OECDs and STEP (Skill Toward Employment and Productivity) programme of the World Bank for the non-OECDs. The average reading proficiency in Plausible Values (PVs) for the OECDs is 277 from the 2012 PIAAC. Similarly, as evident from the 2013 STEP data, Kenya's average reading proficiency in PVs is about 178. Ghana and Kenya have the least average reading proficiency and least average educational attainment relative to other participating non-OECDs in the World Bank's STEP programme. See Figures 2.1-2.3 for score distributions of typical OECDs and non-OECDs.

[20] In this study skill is taken to mean cognitive skill proxied by reading proficiency.

[21] For the few studies in developing countries, see Case and Deaton (1999) Pritchett (2001); Glewwe and Kremer (2006); Hanushek and WoBmann (2007); Glewwe and Miguel (2008).

schooling or issues of insufficient school inputs deemed to characterise schooling in developing countries. The ideas of these existing studies are in concordance with the study of the Independent Evaluation Group (2006) of the World Bank (see article, 'From Schooling Access to Learning Outcomes: An Unfinished Agenda' (Nielsen, 2006)) that gives insights on how the objectives of school provision in developing countries have evolved. The World Bank's study suggests that, over time the emphasis has shifted from mere enrolment; to enrolment and completion; and of late, the focus has been on raising learning outcomes through school inputs (or a focus on quality of schooling). Hence, of late the emphasis is on raising the quality of schooling, as a way of raising skills from schooling. The study further suggests focusing on the 'disadvantaged' to eradicate poverty in developing countries. However, it acknowledges the high unit costs to achieve such. This World Bank's study and the related mainstream studies did not only disregard the effect of variability[22] in schooling on the variability in skill across developed and developing countries, but they also failed to acknowledge some key peculiarities of developing countries. It is important to note that factors such as the effects of quality of schooling, background characteristics and peer or neighbourhood effects may vary substantially between developed and developing countries, these substantially impact educational attainment and hence, the skills acquired, across regions of the world. However, the mainstream literature, particularly, the strand that focuses on developing contexts has paid little attention to the effects of background, peer or neighbourhood characteristics, merely emphasising quality of schooling. Hence, in this study, in examining the extent to which schooling impacts reading proficiency, in developing contexts, beyond the emphasis on quality of schooling, I focus on background characteristics as crucial mechanisms through which schooling impacts skills. I will now discuss these factors[23] in light of the argument[24] raised. Rather than an attempt to argue against the approaches of the mainstream literature, this study presents new and more robust perspectives on examining skill proliferation, particularly, for sub-Sahara Africa, with evidence from urban Kenya. In a subsequent chapter (chapter 4) on externalities of schooling and skill, I consider the effects of aggregate schooling and district size (number of households) as measures of neighbourhood characteristics that influence skill proliferation.

[22] These mainstream studies (strongly) assume schooling give little or no skill in developing contexts.

[23] These factors include quality of schooling, and background characteristics.

[24] The argument (descriptive evidence) suggests, regardless of the quality of schooling, variability in schooling has a clear association with variability in skill, across countries of the world.

Raising the quality of schooling may raise skills as most studies suggest. However, in this study, I argue that the variability in 'quality or school inputs'[25] is at best, only in part, a factor that explains the variability of skills from schooling. This argument is supported by the findings of Hanushek (2003). The study of Hanushek (2003) provides a comprehensive examination of the effects of school inputs and asserts that school inputs have minimal impact on achievements or learning outcomes. However, Hanushek (2003) distinguishes teacher effectiveness as useful school (quality) input that impacts learning outcomes. Furthermore, Hanushek and Luque (2003) find that the notion that school inputs are more important in developing contexts is not supported, as the empirical evidence suggests such inputs have little or no effects on learning outcomes, particularly, in developing contexts. Hence, put together, Hanushek (2003); and Hanushek & Luque (2003) disagree with the mainstream literature in the developing contexts that advocate school quality inputs for skill. However, whilst the conclusions of Hanushek (2003); and Hanushek & Luque (2003) on the effects of school inputs did not specifically argue for or against the main argument of this study that suggests, variation in schooling explains variation in skills, regardless of school inputs or quality of education, what is made clear by Hanushek (2003); and Hanushek & Luque (2003) is that, besides teacher quality, the effects of all other school inputs have no material effects on skills.

The emphasis on the effects of background characteristics on skill in this study and the variability of schooling on skill as argued earlier, motivate the argument for an examination of equity in access to schooling, relative to quality (or school) inputs for skill. It is important to note that raising the quality of schooling where resource constraints prevail can mean access to such 'quality schooling' may be inhibited, as affordability (or background) may impact access for the disadvantaged, relative to the advantaged. Suggesting, raising the quality of schooling may mean lowering access to schooling. On one hand, examining the role of background characteristics and access to schooling on skill gives a useful policy understanding of the dependence of educational attainment and skill on background characteristics. This understanding can influence reforms that impact investment in human capital. On the other hand, the role of school inputs such as class size and teacher characteristics on student performance is useful in understanding strategies on resource utilisation in schools, particularly, this can be directed towards influencing the quality of the labour force. The former – effects of background and access – is the focus of this study as the latter – the role of school (quality) inputs on skill –has been explored extensively. Family

[25] Here, school inputs make measures of quality of schooling.

background characteristics are known to be important predictors of educational (and labour market) outcomes in the OECDs, particularly, in the United States of America (see Hanushek and Luque (2003)). However, no single-country study in a developing context has explored extensively, the effects of schooling on reading proficiency (skill), with emphasis on the mediating effects of background characteristics, to influence policies on educational finance or investment in human capital. To emphasise this again, rather than oppose these existing mainstream studies (see Case and Deaton (1999); Pritchett (2001); Glewwe and Kremer (2006); Hanushek and WoBmann (2007)) that emphasise quality inputs as a way of raising learning outcomes and by extension, labour market outcomes. This study attempts to improve understanding of the role of background characteristics[26] on educational attainment, and in turn, on skill (regardless of school input). Interestingly, the modelling approach adopted and the emphasis on background characteristics as mechanisms through which schooling raises skills is partly, complementary to the existing mainstream studies that emphasise quality in schooling for skill. Particularly, with the 'years of schooling' as input in the models used in this analysis, and at the same time having 'years of schooling' as an outcome/output in the educational production model for quality inputs, makes the models of the skill production function in this analysis (that account for background characteristics) to sufficiently account for 'all existing school or quality inputs' that can impact skill outcome. Hence, the modelling approach of this study is not only sufficiently robust but complementary to approaches of the mainstream studies.

Another justification for the emphasis on background characteristics (and hence, access to schooling) for skills in this study is the evidence of the high proportion of adults and children with no formal schooling in sub-Sahara Africa and the relatively low educational attainment in the region. This suggests the substantial heterogeneity in schooling and skill may be attributable to background characteristics. Hence, the main rationale for the emphasis on 'background characteristics for access' over 'school or quality inputs' is owing to the peculiarities[27] of the developing contexts, particularly, sub-Sahara Africa where limited resources for public provision of schooling abound. Hence, a consideration or efforts to achieve the right balance between 'equity in access'[28] to schooling through background

---

[26] Background characteristics in the forms of parental education and wealth (socioeconomic status of respondents at age 15) can give insights on inequity or inequality in access to schooling and skill.

[27] Most non-OECDs, particularly sub-Saharan Africa are marred by issues of inequality and inequity in schooling and skills. This includes background or socioeconomic characteristics that impact access to schooling and skill acquisition, evident from the substantial percentage of adults without formal schooling.

[28] In this study, with the trend in schooling and skills highlighted, it is argued that efforts towards raising access to (or years of) schooling entails raising resources school inputs that raise capacity (as 'capacity inputs')

characteristics; and quality in schooling, is crucial for optimal skills and labour market outcomes. As discussed, efforts to raise quality in schooling may hamper efforts for equity in access to schooling and vice versa. Hence, with a high proportion of adults with no formal schooling as earlier emphasised, efforts to raise quality over access to schooling may not be ideal for sub-Sahara Africa. The interesting study of Coleman (1966) in the United States of America gives credence to background characteristics and peer or neighbourhood effects for useful schooling outcomes. This further accentuates the argument that school quality inputs are at best of minimal effect on skill, compared to the effects of background characteristics on achievements, learning or labour market outcomes. This brings the arguments of Hanushek, 2003 and those of Hanushek and Luque (2003) that not only suggest consideration for a useful balance between equity in access, and efficiency in schooling but suggest quality inputs have minimal effects on schooling outcomes, however, with an exception to teacher effectiveness. The Coleman comprehensive report (Coleman, 1966) does not directly oppose the ideas of the Independent Evaluation Group (2006) of the World Bank (IEG, 2006)) and the related mainstream studies that emphasise raising skills via quality inputs. The Colman Report is clear about its emphasis on background characteristics and peer or neighbourhood effects, over quality inputs. Put together, these studies (Coleman, 1966; Hanushek, 2003; and Hanushek and Luque, 2003) and the peculiarity of sub-Sahara Africa strongly inspire the consideration for the direct effects of background characteristics—proxied by parental education and wealth—on schooling and skill, in this study.

In summary, motivated by the current trend in schooling and skill across regions of the world, as earlier discussed, this study entails a close examination of the effects of schooling on skills, in a non-OECD context where a good proportion of the adult population is without formal schooling. Moreover, whilst access to schooling remains crucial for mitigating the adverse effects of background characteristics on adult outcomes (Schiefelbein and Farrell,1984), the emphasis of the Coleman Report (1966) and the peculiarity of the non-OECDs strongly suggest background characteristics are crucial for a good understanding of

for more places, this is deemed to raise learning outcomes (or skills from schooling). In this study, the use of 'capacity inputs' differ from 'school inputs' directed towards raising quality of schooling (as 'quality inputs'), conventionally 'quality inputs' are deemed to raise learning outcomes or raising skills from schooling. Limited available resource means a useful resource allocation between 'capacity' or 'quality' of schooling is crucial in maximising schooling outcomes (such as skill acquisition) in line with the objective or need of the context. In this study, **Equity in Access** as used in this study, entails efforts to equitably raise access to schooling or educational attainment (hence, average years of schooling). Hence, this may entail efforts directed towards minimising the percentage of those with no formal schooling, possibly, by impacting background characteristics and this can entail the use of reforms. In this study, background characteristics include parental education or socioeconomic status.

the relationship between schooling and skill. Put together, these suggest, a possible bidirectional relationship between schooling (or skill) and background characteristics. Several studies in the OECDs acknowledge the role of background characteristics in explaining adult outcomes (Green et al., 2015; Almond et al., 2018). However limited evidence for the non-OECDs exists. The studies of Glewwe (2013); and Ravallion (2014) not only suggest the lack of such empirical evidence in the non-OECDs but also suggest the effects of background characteristics on adult outcomes for the non-OECDs, particularly, sub-Sahara Africa is of policy concern. Hence, in this study, beyond the objective to examine the extent to which schooling impacts learning outcomes – in the form of adult reading proficiency – this study considers comprehensively, the effects of background characteristics as mechanisms through which schooling impacts skill in a developing context. Whilst most of the studies considered so far are empirical analyses at the school-age levels, this study focuses on adults who constitute the labour force. Hence, this study examines the mediation of background characteristics in the effects of (years of) schooling on adult reading proficiency (skill). These give useful insights into the effects of inequity and inequality in access to schooling and more importantly, the efficiency of schooling in urban Kenya.

The adult skill profile of Japan and that of two countries in sub-Sahara Africa reveal:

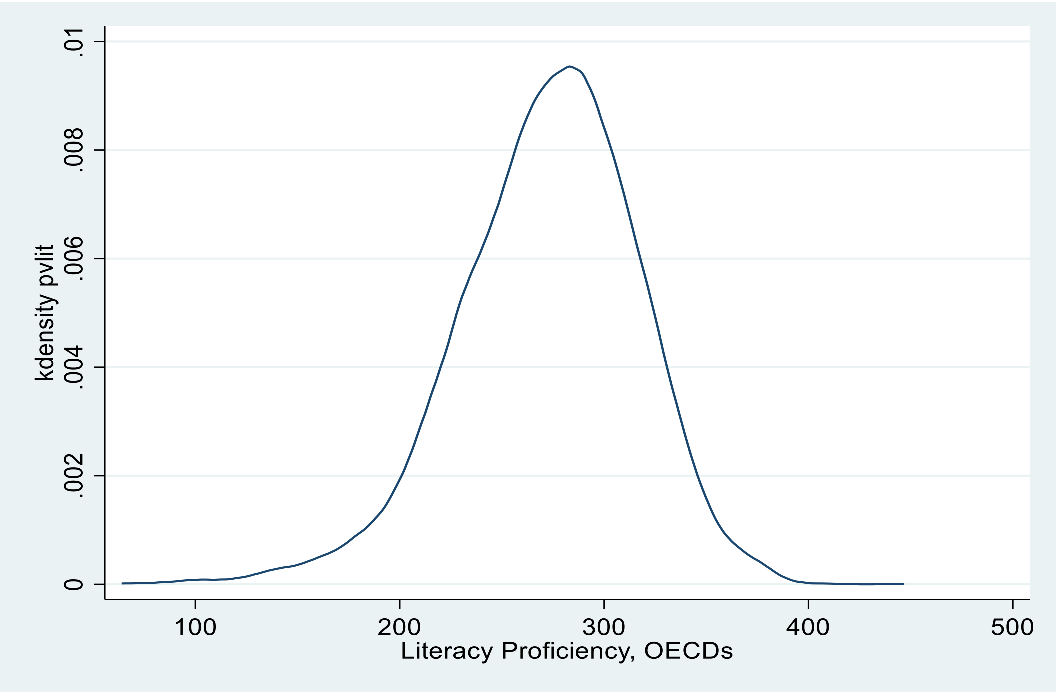


Figure 2-1 Kernel Density Plot, PIAAC (test scores) for an OECD Country (Japan).
Notes: Figure 2.1 is a kernel density plot, showing typical distribution of Plausible Values (test scores) of reading proficiency) in Japan (sample is representative of the labour force) – Data Source: Programme for the International Assessment of Adult Competencies (PIAAC) of the OECD.

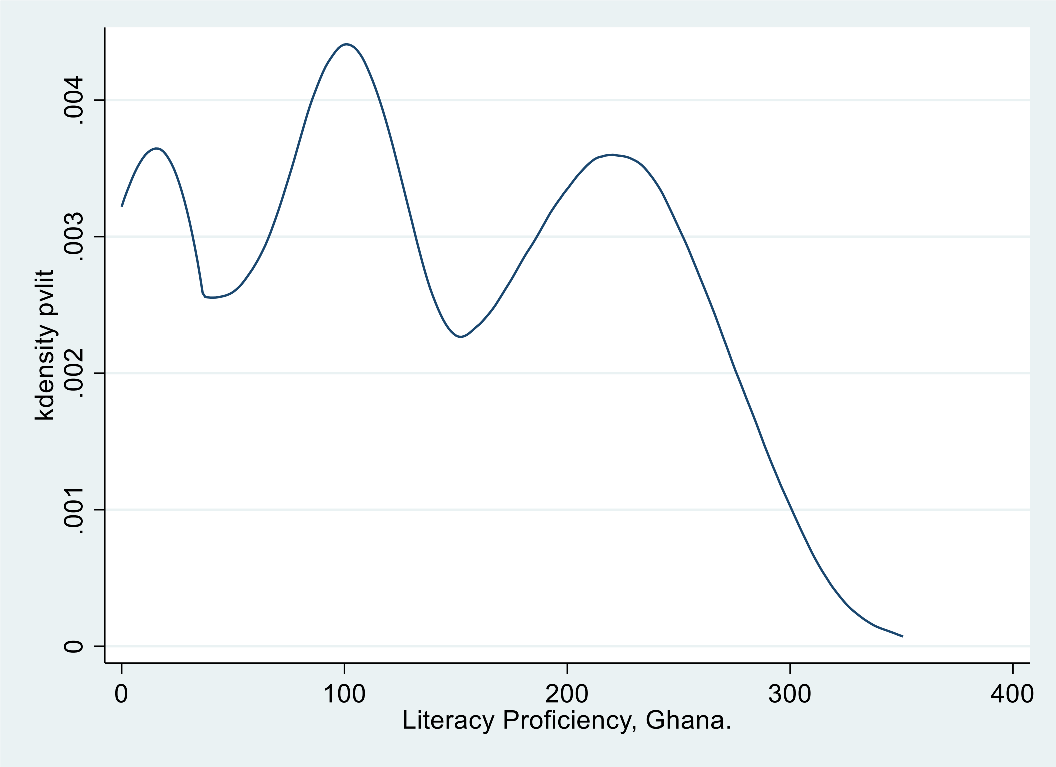


Figure 2-2 Kernel Density Plot, STEP (test scores) for a non-OECD Country (Ghana).
Notes: Figure 2.2, is a kernel density plot, showing the distribution of Plausible Values (test scores) of reading proficiency in Urban Ghana (sample is representative of the labour force) – Data Source: STEP Skills Measurement Program (in low- and mid-income countries) of the World Bank.

Figures 2.1, 2.2, and 2.3 present the kernel density plots showing the distribution of cognitive skill (proxied by reading proficiency) for Japan (this shows the typical skill distribution for the OECDs based on PIAAC), Ghana, and Kenya (typical skill distribution for the non-OECDs, based on STEP) respectively. The distributions give useful insights beyond the mean Plausible Values (see the Data Section for PV methodology) of — 297, 135, and 179 — for Japan, Ghana, and Kenya respectively. The distribution of Japan has a single peak (unimodal). However, Ghana has three peaks and Kenya has double peaks. These multiple peaks (multimodal distributions) in Ghana and Kenya, relative to the typical distribution in the OECDs suggest substantial differences in skill and skill formation between the OECDs and sub-Sahara Africa. With the bimodal distribution in Kenya, factors that can potentially explain such structural differences in skill across the distribution include, the impact of substantial changes in schooling reforms which impacts schooling or educational attainment across generations. Overall, the data points (from origin) of the peaks for Ghana and Kenya show that, relative to Ghana, the Kenyan labour force has higher reading proficiency. This is clear as the higher peaks of the distribution for Ghana are fatter towards the left tail of the distribution but the higher peaks for Kenya are fatter towards the right tail of the curve. Overall, these distributions show variability in skill between the OECDs and the non-OECDs and further suggest substantial differences in schooling and background characteristics, across the OECDs and the non-OECDs. These stark differences in the distributions of the OECD country relative to the non-OECD countries suggest the need to investigate the role of inequality/inequity, particularly as it impacts schooling in the non-OECDs as earlier argued (see motivation and literature review).

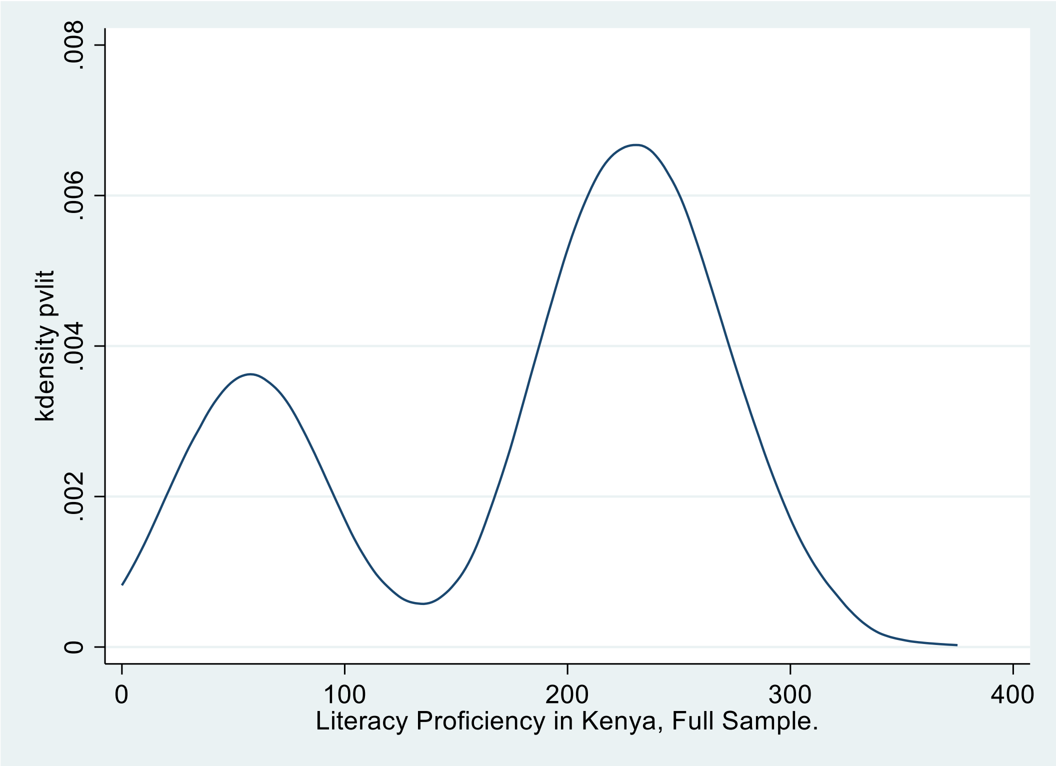


Figure 2-3 Kernel Density Plots, STEP (test scores) for Urban Kenya.
Notes: Figure 2.3 is the Kernel density plot, showing the distribution of Plausible Values (test scores) of reading proficiency, in Urban Kenya (sample is representative of the labour force) – Source: STEP Skills Measurement Program (in low- and mid-income countries) of the World Bank.

In light of this anomaly in the non-OECDs, I now turn to state the research questions that entail examining the extent to which access to schooling impacts skills from two perspectives: firstly, indirectly, by an examination of background characteristics (and related reforms), particularly, how background characteristics impact schooling and in turn, skill (1). Secondly, I examine how schooling in urban Kenya impacts skills, fully accounting for background characteristics (2). Hence, the overarching objective of this chapter is to understand the extent to which schooling impacts skill (2), accounting for the mediating effects of background characteristics argued to substantially explain schooling, skills, and effectively, act as mechanisms through which schooling impacts skill (1).

For emphasis, this subsection presents the objectives (research questions) of this chapter as follows:

(1) What is the extent to which background characteristics (and related reform) impact schooling, and hence, skills in urban Kenya?

(2) What is the extent to which schooling (and related reforms) impacts skills in urban Kenya?

All models account for peer (or neighbourhood) effects, by controlling for average schooling, across districts; and district size (number of households).

2.1.2 Related Literature, Theoretical Framework, Kenya and the 1985 Curriculum Reform

This review is a blend of two distinct but interconnected economic literature—the Economics of Education and Labour Economics—specifically, this study contributes to human capital development, in a developing context, particularly, contributing to the Analysis of Education (I21); Education Finance (I22); Education and Inequality (I24); and Government Policy (I28) that also impacts employment. In this study, I exploit variations in schooling and skill attributable to the 1985 reform. Hence, for ease of exposition, after an extensive review of the main literature, I discuss the 1985 curriculum reform in Kenya, which sets the scene for an analysis of the reform-affected return estimates in this study.

2.1.2.1 Antecedents: Related Literature and Framework

This study is a blend of two interesting literature. The first relates to the first research question on the effects of background characteristics on schooling and skill. This relates to the literature on intergenerational mobility and inequality (for a review, see Black and Devereux, 2011; and Solon, 1999), improving understanding of socioeconomic mobility between parents and their wards. The second literature relates to the second research question that entails examining the efficiency of schooling—in this study, this is the extent to which schooling impacts skill—accounting for background characteristics. This study relates to the literature on the Analysis of Education (see Hanushek, 1970; 1986; 2010; 2015; and 2020). Comparing the mainstream studies[29] that focus on the effects of school inputs on schooling (and skill) to this study that focuses on the impact of background characteristics on schooling and skill (and in turn, the effect of schooling on skill) provides a basis for an argument for useful education financing. Effective education financing is of central

[29] This includes studies in the mainstream literature, such as those of Case and Deaton (1999) Pritchett (2001); Glewwe and Kremer (2006); Hanushek and WoBmann (2007); Glewwe and Miguel (2008).

importance, supporting the goals[30] of education in a context (see, Rice et al., 2020; Guthrie and Wong, 2012; and Monk, 1990). Such goals can include useful skill proliferation from schooling. Specifically, with the emphasis on equity and efficiency for public (education) finance (see Baker and Green, 2009; 2014), this study adds to the literature on Education Finance. Finally, both research questions relate to a strand of the literature on the effects of peers (see Sacerdote, 2011) and neighbourhood (Manski, 1993) on human capital development. Particularly in this study, I have accounted for peer or neighbourhood effects, by controlling for average (or aggregate) schooling on individual skill and earnings. For consistency across this study, all analyses (research questions 1 and 2) account for the effects of average schooling across districts. However, the effect of average (or aggregate) schooling on the skills of individuals is discussed as human capital externalities in a subsequent chapter.

This review of the related literature includes three subsections as thus. The first subsection discusses findings from the related literature regardless of the method or framework followed, emphasising the gaps and how this study fills the gap. Based on the first research question, studies on the effects of background characteristics on schooling, skills, and related outcomes are reviewed in this subsection. Next, the set of studies that consider equity and efficiency in schooling and skill (based on the second research question) are then reviewed, with gaps highlighted, stating how this study fills the gap. To wrap up the first subsection of the Review, I turn to review studies that emphasise the effects of aggregate or average schooling and skills across cities/districts as externalities of human capital that may impact individual schooling and skill. Turning to the second subsection of this Review, here, related studies based on the framework followed are reviewed, this is particularly important for the first research question on intergenerational mobility. Lastly, to set the scene for the reform-affected estimates considered in this study, the final subsection presents a review of the 1985 reform implementation in Kenya, acknowledging the effects on schooling and skills outcomes based on the related studies on reforms (see Downes, 1992; and Karp, 2015) in the OECDs.

[30] Regions where there are substantial proportion of the labour force that have little or no schooling may prioritise more access to schooling for skill, whereas regions where the labour force is highly educated or skilled, then the right step may be to further raise quality or school inputs.

Generally, all reviews start from the wider literature (regardless of the OECD or non-OECD context), this is then narrowed to studies in the non-OECDs, and finally, the few available related studies in Kenya are discussed.

2.1.2.1.1 Reviews of the Related Literature—Some Findings.

Inequality in Schooling and Skill – Effects of Background Characteristics

Despite the widespread understanding of inequality in education and skill (see, Solon 1999; Black and Devereux, 2011), only a few single-country studies have (empirically) examined the effects of background characteristics on educational attainment in sub-Sahara Africa (see, Alesina et.al. (2021); Azomahou and Yitbarek (2016)). Most existing studies in sub-Sahara Africa are cross-national studies with some inherent limitations, as earlier argued, conclusions of cross-national studies typically, either apply to ‘all countries’ as an ‘average’ or do not apply to the specifics of any of the countries. This raises questions as to the true (actionable) policy relevance of the main conclusions from such studies to a given country as the very nature and effects of background characteristics (socioeconomic status and parental education), the associated peer effects and related reforms vary substantially across countries even with substantial within-country heterogeneity.

Education is considered a crucial determinant of the opportunities that accrue to individuals in their adulthood (Stiglitz, 2012) and the study of Gabay-Egozi and Yaish (2019) in Israel shows (empirically) that intergenerational educational mobility[31] is associated with striking advantages/disadvantages over life. A meta-analysis by Hertz et al., 2007 reviewed intergenerational educational mobility spanning over fifty years for 42 countries, across both the OECDs and the non-OECDs, inclusive of sub-Sahara Africa (Ethiopia, Ghana, and South Africa). Findings suggest that the elasticity of a child’s education, relative to their parent’s education is in the region of 0.3 - 0.5. Further findings suggest that the regression coefficient

[31] Educational (intergenerational) mobility is simply a term used to describe movements in the educational attainment of parents and their children (in their adulthood) over time. This may be persistent (no substantial movement), upward or downward mobility (movements). Educational mobility is a crucial tool for economic growth as it can be a useful determinant of social and economic mobility which, over time, can either raise individuals/households out of poverty on make them poorer. Mobility (intergenerational) in education (and skill) from parents to offspring impacts social and economic mobility, which over time, strongly influence the ‘income status’ of individuals and in aggregate, countries they live in. Kenya is considered a low-income country and plans to be a mid-income country by 2030, hence its ambitious plans for education (education is considered a crucial vehicle for such growth plans). Furthermore, patterns between the socioeconomic status (inclusive of education and earnings) of parents and the outcomes (education and earning) of their wards give useful insights into the level of inequality of opportunities across the sub-populations of a country.

(elasticity) has fallen over time, suggesting lower persistence. Hence, over time, a higher mobility between parents and their offspring. However, the study of Hertz et al., 2007 finds that the correlation coefficient (a measure of the observed dispersion of the offspring's schooling that is explained by parental education) remains unchanged over time. The interesting study of Hertz et al, (2007) cuts across the OECDs and the non-OECDs, findings suggest substantial heterogeneity in intergenerational educational mobility across regions, with the least mobility in Latin America and the greatest mobility across the Nordic regions. Although the findings suggest Latin America has the greatest persistence relative to other world regions, this does not necessarily suggest countries in sub-Sahara Africa involved (Ghana, South Africa, and Ethiopia) have greater mobility as a possible reason for their lower persistence may be due to the minimal average schooling in sub-Sahara Africa, relative to Latin America.

Turning to studies in the non-OECD contexts including specific studies in sub-Sahara Africa, the recent work of Alesina et.al. (2021) sets the scene for the related reviews in this region. Alesina et.al., (2021) examine intergenerational mobility in education for over 2800 regions, in over 360 provinces across 27 countries that span north, south, west, east, and central Africa inclusive of Kenya, with the 1969, 1989, 1999, and 2009 censuses used. Findings are interesting but considered 'generalised', hence may not relate to any specific country or region considered. However, the study of Azomahou and Yitbarek (2016) focuses on fewer (9) countries in sub-Sahara Africa, without Kenya. Alesina et.al., (2021) acknowledge wide cross-country (and within-country) heterogeneity in intergenerational mobility with substantial heterogeneity across rural-urban settings. Alesina et.al., (2021) affirms the evidence of a significant gender gap in intergenerational mobility in some countries but affirms this is not 'systematic' across the countries. Again, the study of Alesina et.al., (2021) provides (and supports) a basis for further investigations into a single country (region), for more 'specific' findings that can drive useful policymaking at the country/regional levels.

This study for Kenya relates to Alesina et.al., (2021); and Azomahou and Yitbarek (2016), providing similar findings in the context of urban Kenya. Although the study of Azomahou and Yitbarek (2016) found less heterogeneity across the countries, relative to Alesina et.al., (2021) this may be attributable to the common political events that impact most of the countries in sub-Sahara Africa, pre- and post-independence. However, 'finding less heterogeneity' is not the absence of heterogeneity, as intergenerational persistence or mobility across countries may be driven by characteristics peculiar to a given country/region, suggesting certain variables/factors may remain unobservable in a cross-country context.

Suggesting a single-country study where specific characteristics that define and improve estimates are fully observed. Alesina et al. (2021) examine intergenerational educational mobility across twenty-seven (27) African countries. Absolute upward mobility is defined by the likelihood that children born to parents without primary education manage to complete primary schooling. Absolute downward mobility is defined by the likelihood that children born to parents who completed primary schooling do not manage to complete their primary education. The findings from Alesina et al. (2021) not only suggest variations in upward and downward (intergenerational) educational mobility but within-country variations, across the twenty-seven (27) African countries. Specifically, Alesina et al. (2021) find that, whilst South Africa and Botswana are most likely (over 70% chance) to have upward mobility; and Ethiopia, Sudan, Mozambique, Guinea, Malawi, and Burkina Faso are least likely (less than 20% chance) to have upward mobility. Alesina et al. (2021) also find that the literacy proficiency of the old generation explains about fifty percent (50%) of the observed disparities in intergenerational educational mobility across Africa.

Further findings from Alesina et al. (2021) show that, for a country like Kenya with an average of 50% chance for upward intergenerational educational mobility, substantial within-country variation is evident in Kenya. Specifically, in Kenya, the Turkana region only has about 5% likelihood of upward mobility, whereas up to 85% likelihood of upward mobility exists for the Westlands in Nairobi. This suggests besides parental education, skills, and other socioeconomic characteristics of parents, other factors such as peer effects (or environmental factors) may be crucial drivers of mobility. This study relates to Alesina et al. (2021) who emphasised how geographical and historical variables across post-independence Africa impact upward and downward intergenerational educational mobility. Suggesting the geographical and historical variables linked to regional development may be ‘additional’ mechanisms through which peer effects or average schooling and skills across districts impact individual schooling and skill. However, with a focus on Kenya, this study extends this knowledge by examining the extent to which parental education (and wealth) impact not merely the schooling, but the skills of the offspring, accounting for average or aggregate schooling and skills across districts as possible externalities of human capital. In this study, I have controlled for average or district-level schooling and skills in determining the effects of parental education (and wealth) on the schooling and skill of individuals. However, in a subsequent chapter – Chapter 4 – a comprehensive analysis of the effects of average (or aggregate) or district-level schooling as externalities of human capital, and as they impact skills (and earnings) in urban Kenya are discussed.

A country-specific study would unravel the extent to which background characteristics and the associated reforms explain years of schooling and skill in an intergenerational framework. As earlier argued, findings from such country-specific studies provide more policy-relevant or actionable conclusions compared to cross-country studies. Furthermore, unlike the non-OECDs, several recent single-country studies are available in the OECDs. These include the study of Checchi et al. (2013) for Italy and those of Cobb-Clark & Nguyen (2010) for Australia. The former (Checchi et al., 2013) finds a high persistence in educational attainment in Italy, particularly, a high probability of obtaining a university degree for children of highly educated fathers. The latter (Cobb-Clark & Nguyen (2010)), finds, that immigrants from families with non-English speaking backgrounds have an educational advantage over their peers (immigrants with English-speaking backgrounds and the Australian-born). Interestingly, Cobb-Clark & Nguyen (2010) find the education of young Australian immigrants (with non-English speaking backgrounds) is most impacted by the education of their mothers, whereas the education of those with English-speaking backgrounds is most influenced by the education of their fathers, like the case of Italy (see Checchi et al. (2013)). However, unlike the case of Italy, highly educated Australian-born parents transfer roughly equal (but separate) educational advantages to their wards. Cobb-Clark & Nguyen (2010) further suggest that the intergenerational mobility of families of Australian-born; and immigrants with highly educated parents of English-speaking backgrounds is much the same. Lastly, intergenerational educational mobility is much greater for families with non-English-speaking backgrounds.

These findings for single-country studies bear useful policy relevance for educational attainment. However, is there an existing single-country study for Kenya? These single-country studies in the OECDs relate to this study as they examine how the mother and father's schooling impact the schooling of offspring, beyond that, this study, examines how parental education impacts the skill of the offspring, accounting for peer effects in the form of average (or aggregate) schooling across districts. As earlier highlighted in the empirical literature on intergenerational educational mobility, relative to other regions of the world, sub-Sahara Africa is only beginning to gain some attention, with relatively few single-country studies that are mainly concentrated in South Africa[32]. These interesting studies for South Africa suggest persistence (in educational mobility) especially, among the poor and black race. However, more recent studies now find, rising educational mobility in South

[32] See the related study of Kwenda et al. (2015); Branson et al. (2012); and Nimubona and Vencatachellum (2007).

Africa. This change may be due to schooling reforms (see Jansen, 1998) in South Africa. This is the first single-country study for urban Kenya that adds to the economic literature on intergenerational educational mobility between parents and their offspring. Filling this gap for Kenya means similar estimates as those of the OECDs and other non-OECDs become available for Kenya.

Efficiency (Quality) in Schooling — Effects of Years of Schooling on Skill

The following review is on the related literature on the 'efficiency' of schooling. Efficiency in this context, is a term used to describe the effect of schooling (hence, years of schooling) on skill (reading proficiency) or simply, the extent to which schooling impacts reading proficiency, as used in this study. For a few existing studies in sub-Sahara Africa, see Shafiq & Valerio (2021); and Lucas & Mbiti (2012). Most existing related studies are for the OECDs (for a few, see Lee and Wie (2017); Aakvik et al. (2010); Dee (2004); and Ganzach (2000)). In examining skills from schooling in an intergenerational framework, this study relates to the study of Lee and Wie (2017) that examines skills from schooling in Japan and South Korea and finds, relative to the mother's education, the effects of the father's tertiary education on skill (literacy, problem-solving, and numeracy) is substantial. However, findings further suggest that, relative to background characteristics, schooling has a higher mean and statistically significant effect on skill. Although this evidence suggests schooling (significantly) explains skill across both countries, further findings affirm schooling is more efficient in Japan relative to Korea, as higher levels of schooling explain the skill of numeracy and literacy in Japan, and this is not exactly the case for Korea where effects are statistically insignificant. This study presents similar evidence in urban Kenya, improving understanding of the efficiency of schooling in a country in sub-Sahara Africa with its distinct (and characteristic) skill profile, relative to the OECDs (see figures 2.1-2.3). The study of Shafiq and Valerio (2021) examines the effects of several family background characteristics and schooling on skill for eight (8) developing countries inclusive of Kenya and Ghana in sub-Sahara Africa, their study provides useful descriptive evidence. They found that relative to parental socioeconomic status (inclusive of parental education), the schooling of wards explains their skill. In this study, using the STEP data for Kenya, I examine the effects of background characteristics and schooling on skill (reading proficiency). Beyond this, also examined is how the different categories of schooling or educational attainment of respondents impact their skill—this gives useful insights into the efficiency of schooling. This study examines the claims of Pritchett (2001) that suggest schooling yields little or no skill, following the approach of Lee and Wie (2017), I examine this claim in the context of Kenya by taking a step further, not only accounting for the effects of parental socioeconomic status (including parental schooling) but also aggregate schooling. Doing this presents more useful insights on the efficiency of schooling and related reforms on skill, improving understanding of Pritchett's (2001) propositions in the context of urban Kenya. Lucas and Mbiti (2012) is a related single-country study for Kenya. Without 'directly' accounting for parental education, Lucas and Mbiti (2012) show empirically that free primary education in Kenya raised completion rates in primary education, without any significant reduction in the performance in the exit examination scores. Consistent with earlier arguments, this suggests quality is not significantly affected by free education intervention that raises access to schooling for the disadvantaged. The work of Lucas and Mbiti (2012) shows that the quality of schooling is considerably sustained with the intervention for increased access to schooling for the disadvantaged (with low socioeconomic status). This evidence[33] strongly suggests that carefully planned interventions with a good understanding of background characteristics can result in a substantial rise in schooling, raising average schooling and skill (with little or no adverse effects on quality of schooling) by mitigating the adverse effects of background characteristics. This study differs from that of Lucas and Mbiti (2012) as it uses a newer dataset and goes beyond primary schooling. Using a newer dataset, the effects of (categories of) schooling on skill give a more holistic approach to examining the effects of schooling on skill, fully accounting for background characteristics which include parental education and wealth.

2.1.2.1.2 Reviews of the Literature—Theoretical Framework

There are numerous approaches to conceptualising educational attainment by Economists and Social Scientists, this includes quantitative or qualitative measures such as years of schooling and skills. Measures of 'cognitive' skills include proficiency in reading, writing or problem-solving. A review of the economic literature shows three approaches to modelling educational attainment. This can range from educational production models (see Hanushek, 1986; and Hedges et al., 1994) where the 'outcome' variable is typically, a qualitative measure of schooling such as test score, expressed as a function of certain 'inputs' (or independent variables) that may range from school (and family) characteristics. In this framework, the variability of the outcome comes with the variability of the 'inputs' and technology. Here, educational attainment is not impacted by the choice of the individual. In

[33] As earlier argued, at best, quality inputs, partially explain skills. This stance is well supported by the study of Lucas and Mbiti (2012) in urban Kenya. Suggesting useful effort to raise skill equitably must entail provisions for equity in access to schooling rather than increased school inputs that at may adversely impact equity in access especially in times of resource constraints.

contrast to the educational production framework is the human capital framework (see Becker (1964)) where educational attainment is modelled based on the well-informed choice of the rational individual with an understanding of the expected returns to schooling (see Freeman (1975; and 1986)). This choice can come through the decisions, nature, and nurture of parents (see Becker and Tomes (1986)). A third approach explored in the empirical literature is somewhat a hybrid of the educational production and human capital framework, with 'any' well-defined input and outcome variables. However, structures (assumptions that may be complex) may be specified but relationships can simply be examined using reduced-form (without complex assumptions) modelling approaches (see Haveman and Wolfe, 1995). With the research questions, ideas or relationships in this chapter where measures of human capital such as 'years of schooling' and 'skill' make outcomes that are jointly modelled with inputs that include background characteristics (and the related reforms), the flexibility of the reduced-form models means the third approach provides the required support for the modelling in this study. Particularly, the modelling approach of the educational production function may involve specific school inputs (such as measures teacher effectiveness) that are also not fully in agreement with human capital framework[34] may not the appropriate based on the relationships of this study. Specifically, in this study, the approach to modelling educational attainment is mostly inspired by the comprehensive economic perspective proposed by Haveman and Wolfe (1995) where choices of the society (e.g., via reforms), parents (via background characteristics) and the child are emphasised as crucial determinants of the attainment (the schooling and skill) of the child. This objective of examining the productivity[35] of schooling by first assessing the loss or gain in years of schooling and skill attributable to background characteristics and related reform present a novel approach to examining returns to (or productivity of) schooling. Particularly, the consideration for background characteristics and the related reforms for human capital development make this study a unique blend of the literature on returns to educational attainment following the intergenerational mobility framework (see Bjorklund and Salvanes, 2011) and comprehensive economic perspective (see Haveman and Wolfe, 1995) as earlier mentioned. With both studies referencing the seminal work of Becker and Tomes (1985) that emphasise crucial mechanisms of the human capital framework as the 'intergenerational effects' of these background characteristics. Hence, although not entirely new to the

[34] In the human capital framework, educational attainment is modelled based on the well-informed choice of the rational individual with an understanding of the expected returns to schooling.

[35] Whilst studies may use '*productivity of' and 'returns to' schooling* interchangeably, in this study, I use 'productivity of or returns to' schooling to mean an assessment of skill from schooling; 'returns to' schooling may also mean an assessment of earnings from, or the wage effect of schooling as used in subsequent chapters.

economic literature, understanding how schooling and skill are determined via the joint effects of the intergenerational transmission mechanisms (background characteristics) and related reforms need further elaboration. Please, see the Methods and Data Section where the main mechanisms or channels of the intergenerational transmission is modelled.

2.1.2.2 Further Antecedents

This subsection presents further reviews of this study. Specifically, this subsection is made up of two main units. The first unit (subsection 2.1.2.2.1) discusses the key predictor variables (socioeconomic status and parental education), the institutional background and the 1985 curriculum reform. The second unit (subsection 2.1.2.2.2) reviews the literature on the effects of the reform on education and skills.

2.1.2.2.1 Key Variables, Institutional Background, and the 1985 Curriculum Reform

Key Variables:

The mechanisms (Socioeconomic Status and Parental Education) through which 'Quality in Schooling' and 'Equity in Access' impact Skill

Lebeau and Oanda (2020)[36] point out in their study, 'Higher Education Expansion and Social Inequalities in sub-Sahara Africa', that over the past three decades, participation in Higher Education has risen (also see Cohen and Soto, 2007; Barro and Lee, 2013). However, they pointed out that enrolment rates remain modest, relative to the OECDs. At first, this was attributed to the minimal rise in the number of tertiary institutions that did not keep pace with the demand in the region. This is likely due to the high population growth rate in most countries of sub-Saharan Africa. Subsequently, Lebeau and Oanda (2020) raised concerns over issues such as inequality and inequity in access to (higher) education, attributing access to higher education to some social or background characteristics that distinguish the 'elites' from the 'new group'. According to Oanda and Jowi (2012), the action of converting low-fee-paying mid-level TVET colleges to fee-paying universities meant the loss of schooling opportunities as this would mean the loss of places in the relatively affordable TVET colleges and an increase in places in 'high fee-paying' tertiary (university) education. This suggests inequity or inequality in access to university education for those with minimal financial resources. Suggesting government efforts in the forms of reforms and other growth

[36] Of the United Nations Research Institute for Social Development.

strategies may have adverse effects, particularly, at best, resulting in increased educational attainment that is inequitably distributed. Hence, a case where, based on certain characteristics, only a few ‘advantaged’ get (highly) educated and the many ‘disadvantaged’ get little or no education. The study of Ngware et al. (2006) found crucial determinants of access to secondary education in Kenya to include the level of education of the head of the household and the household income among other factors. Several other studies consider health conditions, regional insecurity, and gender imbalances (Achoka et al., 2007). What is clear from most of these studies is that poverty is highlighted as a major factor inhibiting access to education in Kenya, also argued by Achoka et al. (2007). Strongly correlated to poverty are factors earlier emphasised such as parental schooling and wealth at age 15 (socioeconomic status as used in this study) which make crucial background characteristics for schooling and skills. These socioeconomic factors or ‘background characteristics’ as used in this study are emphasised by the reviews of Krueger (2002); and Carneiro and Heckman (2003). The study of Carneiro and Heckman (2003) also emphasised the importance of background characteristics for schooling and skills, particularly, early in life. It is important to note that these studies (Krueger, 2002; Carneiro and Heckman, 2003) are both in the OECD context. In this study, the consideration for socioeconomic status (parental wealth), rather than current parental income depicts a (more fixed or long-term) measure of family welfare that impacts the ward/respondent at their key developmental stage. The measure for socioeconomic status used in this study captures on a 10-point scale, the wealth of the respondent’s family at age 15 (see Data Section). Also of interest in this study is the effect of parental education – with the mother and father’s education examined separately[37]. Although a few studies categorise parental wealth or income; and parental job or education as different measures of ‘socioeconomic status’. However, in this study, besides examining the effects of parental welfare (as socioeconomic status at 15 of the ward) on the schooling and skill of respondents, I seek to examine distinct effects of parental education on the education and skill of the respondents. To this end, I follow a similar approach of Chevalier et al., (2013) that examines parental education and parental income on the schooling of wards. The distinct effects of parental education and socioeconomic status as defined in this review are crucial and worthy of attention as these variables not merely impact access to schooling (and years of education attained) and skill acquired but may do so differently, with strong intergenerational transmission effects strong enough to create systemic inequality or inequity in access to schooling across sub-populations over time.

[37] Please, refer to the data section for variable specifications.

To set the scene for a discussion on the 1985 curriculum reform in urban Kenya. I do this by assessing variations in schooling and skills, based on background characteristics. I highlight the impacts of aid, reforms, and the related expenditure over time as inputs of education and the related outputs in the forms of access and quality of schooling that impact skills.

Institutional Background:

Expenditure, Aid and Reforms as Inputs; Equity in Access and Quality in Schooling as Outputs

Kenya presents a testbed for analysis of the variation in schooling and skill attributable to reforms and background characteristics. Fuelled by reforms, aid and related expenditure, the resultant effects of attempts to create more school places (raising access to schooling); or raise the quality of education (by increasing school inputs) play out in the school expansionary strategy of Kenya. As argued in the previous subsections, these resultant effects of reforms may have some implications on inequality or inequity attributable to the variability of backgrounds of the learners. Drawing from the study of Wycliffe and Colclough (2009) titled: 'Financing Education in Kenya: Expenditures, Outcomes and the Role of International Aid', I now discuss some institutional background of this study.

Education Expansion and Coexistence of Non-Public and Public Funded Institutions

The government of Kenya has continued to expand capacity, increasing school places across all levels of schooling, particularly, for higher education. As of 1970, there were only two universities in Kenya, the public-funded University of Nairobi, and the private-funded United States International University. By the end of 2018, there were sixty-three (63) universities which comprised thirty-one (31) public-funded universities in Kenya (Kenya National Bureau of Statistics, 2019). Like the Further and Higher Education Act of 1992 in the United Kingdom which saw several polytechnics convert to universities. The increase in the number of public universities in Kenya resulted in the conversion of polytechnics and TVET (mid-level) Colleges into universities. This move was criticised as it was deemed to inhibit access to university education for the poor who could not afford tertiary (university) education. Although as part of the plan to widen participation (or increase access to education and skills), the government of Kenya is known to follow a regional approach in distributing educational resources. An example of this is the approach to establishing colleges and universities across regions and districts of Kenya with consideration for a fair distribution. However, this appears to have been insufficient in promoting equality in skill acquisition and schooling opportunities for all, in addition, poverty (deemed to characterise most of the households) as opposed to mere proximity, poses a major constraint to access to schooling. The Higher Education Act of 1992 in the United Kingdom (UK) was enacted when the UK Government provided study loans to all home students willing to study, hence, no serious issue of access to university education was attributable to poverty in the UK following the Act. The rising number of non-public-funded institutions (universities and schools) in Kenya appears to have been bored out of the need to complement the public-funded institutions. Hence, the rising number may have been demand-driven. However, the coexistence of the public and private institutions has neither guaranteed useful access nor quality in education, this is particularly serious for secondary and primary education, in Kenya. Around the world, whilst private education is viewed as a thing of the relatively few who can afford it, this is not exactly the case in Kenya and most sub-Saharan Africa. In most developing countries, education is considered a way to overcome poverty for the informed. However, the stark lack of confidence in public-funded schools results in poor (and informed) households, particularly, those in urban areas, opting for low-cost private (non-state) funded schools. Most of these schools are deemed to have minimal resources that can provide useful access or quality; most operate without government approval[38]. However, it is believed that the relatively few high-cost private schools that the relatively few privileged can afford are well-resourced and can provide useful quality of schooling, this category of privately funded schools is government approved suggesting all minimum requirements for an effectively run school are in place. Peterson and Woessmann (2007); and Machin (2007) assert that expansion in education does not guarantee equity in access to schooling. Hence, despite the unprecedented expansion in primary, secondary and tertiary schooling over the years in Kenya, this seems not to have guaranteed equity in access to schooling, which may have contributed to the current skill level. Beyond quality in schooling, this study argues for increased efforts toward equity in access across all schooling categories. In this study, I examine how differences in background characteristics (parental education; and socioeconomic status) explain differences in schooling and skill, hence, assessing issues of equity in schooling and skills attributable to background characteristics. *An objective of this study is to empirically examine the extent to which effort to expand education (particularly through post-independence reforms) impacts schooling, and then skill, fully accounting for background characteristics (as defined) that are deemed to explain issues of equity in access*

[38] Most private funded primary and secondary schools are not government approved, suggesting they lack the minimum requirements for effectively run schools.

to schooling and skill. I now turn to discuss the implementation of the 1985 curriculum reform in Kenya.

Education Expenditure, Other Inputs and Changes in Outputs—Access and Quality

In some forms, equity in access; and quality of education & training remain major objectives of the government of Kenya which has continued to make efforts, particularly, raising expenditure on education. This effort cannot be solely attributed to the government of Kenya as the recent significant progress in education in Kenya is known to have been driven by several others, especially partnerships involving local communities, civil society, private investors, and other donor organisations. However, increasing enrolment in primary education (as a form of access) has been a major effort of the government of Kenya. Efforts to raise the quality of schooling through inputs, particularly, donations of learning and teaching materials, are known to be mainly the form of effort of other non-governmental (donor) organisations.

Examining trends in enrolment across pre-schools, findings show that pre-school enrolment was 1 218 662 in 1999 and rose to 1 643 646 in 2005. This represents a 35% rise in enrolment between 1999-2005. The number of preschool teachers in 2005 was 72 182 (less than ¾ are trained) representing a 13.4% rise from 2003. It is important to note that pre-schooling is not public funded like primary schooling.

Between 2002 and 2005 an additional 1 500 000 primary school pupils enrolled in Kenyan primary school this was the largest increase representing an annual growth in enrolment of 7.4%. Prior, only an additional 532 400 pupils enrolled at the interval (1996-2002), representing (annual) growth of only 1.5%. This suggests that the Free Primary Education (FPE) programme launched in January 2003 was well received by the disadvantaged in Kenya. Whilst the FPE provided KES 1 020 per child, within these intervals, further evidence suggests a useful rise in GER and NER (Gross and Net Enrolment Rate), the higher GER over NER suggests evidence of over-aged pupils in primary schools. Having overaged pupils in schools is a consequence of several factors. Although evidence shows only a few return or re-enrol, early pregnancies in girls is a major reason for this. Repetition is another related issue that may explain substantial variances between GER and NER, the evidence suggests this is more prevalent in the disadvantaged'. The government now requires automatic[39] promotion in primary schools in Kenya. Progression from primary to secondary schooling is significantly hampered by factors such as poverty or background characteristics. However, the capacity of secondary schools contributes to the issues of access to schooling. Furthermore, it is well known that the availability of textbooks strongly impacts teaching and learning outcomes in developing contexts. Pre-1985, KES 20 (Twenty Kenyan Shillings) per pupil was allocated to provide learning materials, mainly targeted at children from poor households. However, at some points, before the cost-sharing programme, the rise in enrolment constrained the KES 20 provision. The cost-sharing programme transferred the burden of book provision to the parents. This came right after the Kamunge Commission (1988), post-1985. However, again during the 1990/1991 financial year, the supply of textbooks to poor schools recommenced. In 2003, during the Free Primary Education (FPE) programme, 64% of the available funds per pupil (KES 1 020) were allocated to textbooks and other learning materials. The provision of textbooks to pupils in Kenya has changed over the years. With the DFID[40], the World Bank and the Embassy of the Netherlands involved at some points. Over time, the textbook-pupil ratios have been improving. In 1999 the evidence shows a range of 1:5 to 1:10 for lower primary; and 1:2 to 1:5 for upper primary. However, as of 2005, the average was 1:3 for lower and upper primary across key subjects such as Mathematics, English, Kiswahili, Science, Religious Education and Social Studies. As of August 2006, there were 174 576 schoolteachers on duty in 20 229 primary schools.

Unlike primary education, secondary education in Kenya has not always been free. Tuition fees were an average of KES 10 265 per year. However, during 2008 tuition fee waiver was applied in secondary schools. In the ten years to 2006, enrolment rose by about 40% this rise was not attributable to the tuition fee waiver that came later (in 2008) – Affordable Secondary Education (ASE) programme. Amidst the FPE and the ASE programmes, some poor/disadvantaged are still unable to access schooling. This may stem from parental attitude, belief, distance from school, illness, disability or lack of interest. These impact the poor or ‘disadvantaged’ more. Beyond the provision of textbooks, one crucial element of quality inputs is teacher quality, and the Kenyan government has done so much regarding this, particularly through in-service training. By 1994, all untrained teachers were out of the primary schools of Kenya. However, this was not the case in secondary schools, where, as at the mid-2000s, about 13% of teachers at this level had no formal teacher training.

---

[39] Automatic promotion refers to students’ progression across all years, regardless of their academic performance.

[40] The DFID here means the then United Kingdom’s Department for International Development that is currently, the Foreign, Commonwealth and Development Office (FCDO).

However, at this time, most of the teachers were university graduates. The DFID and the British Council among other donors have been very much involved in initiatives aimed at teacher training and management of schools in Kenya. As of August 2006, there were 48,425 secondary schoolteachers on duty in 4 215 secondary schools in Kenya. Overall, performance in the terminal examinations (KCPE and KCSE) for primary and secondary schools shows boys perform better than their female counterparts and private (non-state funded) schools do better than public-funded schools. Within publicly funded schools, those in urban areas do better than those in rural areas. More specifically, from 1990 to 1995, the average score for Mathematics, English, Physics and Chemistry for KCSE was at most 35.7%. Across the same period, the average score in the KCPE for English, Mathematics, Kiswahili and Agriculture is at most 50.26%. These scores show low performances of students across primary and secondary schools, in urban Kenya.

Unlike primary and secondary schooling, Technical and Vocational Education and Training (TVET) has received minimal attention from the Kenyan Government. Between 2002 and 2006, enrolments were quite unstable. Unsubsidised, programme tuition can be quite expensive in polytechnics. Obsolete equipment and abolishing production courses may have contributed significantly to the unstable enrolment. However, over the same years (2002-2006) there have been reforms, particularly, those that expanded technical and vocational education through support from donor countries/organisations. Specifically, reforms play a vital role in the diversification of courses, such as the inclusion of TIVET (Technical, Industrial, Vocational and Entrepreneurship Training) deemed relevant to the needs of the labour market in Kenya. Furthermore, with aid from the Italian Government and affiliations with some existing universities, the Kenya and Mombasa polytechnics were upgraded and expanded to degree-awarding institutions. However, some of these reforms attracted criticism. Ultimately, they have resulted in a useful rise in enrolments. However, the problem of lack of qualified instructors persists. Certainly, unlike university education, the proportion of the ‘disadvantaged’ particularly, female students are more in TVET programmes, although they only make up slightly above 40% of those enrolled in TIVET. Females only make up a third of those enrolled in universities in Kenya. Records of the Teachers Service Commission suggest there were about 3 313 TIVET teachers in Kenya as of August 2006.

Another area of education that has received less attention from the government is Adult Education. Generally, lack of funds has hampered the availability of resources, resulting in a lack of qualified teachers and; a loss of motivation to enrol (by adult learners). Like the TVET in the Ministry of Science and Technology, Adult and Continuing Education is within the Ministry of Culture, Youth and Sport and not in the Ministry of Education, this is deemed a major reason for the current state of Adult and Continuing Education. However joint efforts of the DFID and UNICEF continue to support Adult and Continuing Education in Kenya, particularly, by funding studies that document and report levels of competence, grades, and ages of individuals in Adult and Continuing Education. This has continued to inform policymaking in this area. In 1990, 147 940 adult learners were enrolled, however, this fell 93 052 in 2001. As of 2004, there were about 108 653 adult learners enrolled. About 30% of the learners are males and about 70% are females.

With the increase in private universities and the conversion of polytechnics, there has been substantial growth in University Education, in Kenya. However, relative to enrolments in private-funded universities, more students are in the public-funded universities. Specifically, as of 2004/2005, the number of students enrolled in private universities is 10,050, whereas the number of students enrolled in public universities is 91 541. The percentage of secondary school leavers progressing to university is under 10%. There is a rising demand for university education. With the limited places, evidence suggests that only a third of those qualified have typically gained university admission. Overall, it is understood that university expansion has had minimal support from donor organisations.

On average, expenditure on education as a percentage of total government expenditure pre- and post-1985 is on the rise, from 18.1% in 1980/81 to 18.2% in 1996/97 and 23.7% in 2006/07 with more and more support from the international community. Whilst access to primary education is ‘sufficiently’ equitable, access to secondary, and university education is less equitable. Specifically, only 21.7% of university students are from low-income families as of 2007. As of 2002, the average net enrolment rates (NER) of the poor relative to the non-poor for primary education is 65.0 and 72.7 respectively; for secondary education, it is 6.2 and 17.0 respectively; for tertiary education, it is 3.6 and 4.2 respectively.

The 1985 Curriculum Reform:

Changes in Years of Schooling Pre- and Post-Reform

Post-independence, education in Kenya has continued to undergo transformations supported by several policy initiatives founded by commissions set up to review inputs, outputs, and issue advice. Supported by the Odhiambo Commission in 2017, the Ministry of Education (MoE) through the Kenya Institute of Curriculum Development (KICD) introduced a new competency-based curriculum, which entails a structural reform. This involves a change from the 8-4-4[41] system that kicked off in 1985, to the new 2-3-3-3-3-3[42] structure. This new competency-based curriculum structural reform become active a little over thirty (30) years after the previous structural (1985) reform. The curriculum (structural) reforms and periodic updates can be a useful way to build a first-rate education system whilst improving accessibility to schooling with statutes that can enforce and more importantly, make provisions for compulsory education. Hence, such schooling reforms (and the related periodic updates) can raise access by impacting inputs, capacity, and quality (skill) and effectively raise educational attainment by influencing background characteristics. Typically, this should satisfy the needs of the labour market, achieving useful employment and growth. This study focuses on the labour force, hence, rather than a focus on the 2017 curriculum reform, the focus is on the 1985 curriculum reform that is deemed to have impacted the schooling and skill of the labour force. Following the account of Inyega et al. (2021), I will now turn to motivate this study using the curriculum reforms in Kenya since its independence in 1963.

Education in Kenya has continued to witness more reforms since the country gained independence in 1963. Kenya has recently transitioned to a new curriculum structure. Starting from a 7-4-2-3[43] structure in 1963, to the 8-4-4 structure in 1985, the 1985 curriculum reform added a year to primary schooling, integrated the lower- and upper-secondary school levels, and on average, added a year to tertiary education. The 7-4-2-3 structure was deemed too academically oriented, hence, not meeting the needs of the wider labour market. There was a perception of a lack of flexibility in the structure. This lack of flexibility (attributable to curriculum structure) is argued to result in difficulty adapting to labour market demands. Hence, the 8-4-4 structure was aimed at overcoming the 'undue emphasis' on academics over the supply of skilled manpower required for labour market success. As earlier stated, this change in the curriculum structure resulted in an increase in

[41] (8-4-4) structure entails eight years of primary schooling; four years of secondary and four years of tertiary schooling. This change in 1985 was the basis of the reform dummy used extensively in this study. The 1985 curriculum reform adds a year to the compulsory (primary) schooling, hence, the previous (pre-1985) system required seven years of primary schooling; and the change in 1985 increased the minimum schooling (primary education) to eight years.

[42] (2-3-3-3-3-3) is two years of pre-primary schooling; three years of lower-primary schooling; three years of upper-primary schooling; three years of lower-secondary schooling; three years of upper-secondary schooling; and the number of years of post-secondary schooling differing across programmes, with the certificate programmes of shorter duration relative to university programmes.

[43] 7 years of primary education, 4 years of secondary education, 2 years of high school and 3 years of tertiary education.

minimum schooling (primary education) from seven to eight years, an additional year of schooling.

The latest curriculum re-structure, 2-3-3-3-3-3 launched in 2017 is argued to be 'competency-based'. It includes two (2) years of pre-schooling, six (6) years of primary schooling, three (3) years of lower secondary, and three (3) years of upper secondary schooling. This results in two additional years of education (on average) in the curriculum, in Kenya, excluding tertiary schooling. Primary schooling remains free and compulsory. Although pre-primary education is not mandatory, it is very much subscribed to. Rather than creating entirely new structures, the approach to educational expansion or widening access and institutional growth of the education system prioritises building on existing structures. This includes creating more classrooms within existing schools and further developing existing reforms[44]. In this study, I use the term '1985 curriculum structural reform dummy' extensively. This is deemed to capture the effects of all policy instruments implemented from the beginning of January 1985 when the 1985 structural reform was enforced, to just prior 2017 structural reform. Particularly, the reform dummy (is deemed to) captures the effects of the 1981 Mackey Report which recommended the change from the 7-4-2-3 to the 8-4-4 system. In addition to the Mackey Report, the reform dummy is deemed to capture the effects of the Kamunge (1988) and Koech (1999) Commissions implemented subsequently. The Kamunge Commission recommended cost sharing and strengthening of vocational and technical education, whereas the Koech Commission recommended emphasis on quality and more integrated schooling. Subsequently, the Odhiambo Commission recommended a flexible and comprehensive education structure which was the basis of the change from the 8-4-4 to the 2-6-3-3-3 system launched in 2017. The latter (Odhiambo Commission) is

---

44

Post-Independence					
Ominde Commission	**Gachathi Report**	**Mackay Report**	**Kamunge Commission**	**Koech Commission**	**Odhiambo Commission**
Curriculum Structural Change: 7+4+2+3, Implemented. Emphasis on National Unity; and Responsive Education System. Plan for Racial Integration in Education.	Curriculum Structural Change recommended. Recommends Abolishing School Fees for Primary School	Recommended Curriculum Structural Change from 7+4+2+3 to 8+4+4.	Cost-Sharing Scheme, implemented. Improved Technical and Vocational Education	Totally Integrated Quality Education	Comprehensive and Flexible Structure of Education Recommended.
1964	**1976**	**1981**	**1988**	**1999**	**2011**

deemed not to be captured by the 1985 policy dummy used in this study as it was implemented in 2017, the labour force (those aged 15-64) would be out of school, at this time. Recommendations of the Gachathi Report resulted in abolishing school fees for primary schools pre-1985. Most policy instruments that impact background characteristics may do so indirectly. The curriculum structural reforms that result from the Mackay Report, Ominde and Odhiambo commissions result in changes in years of schooling and are deemed to impact skill. Specifically, the Mackay Report is the basis of the 1985 reform used extensively in this study.

2.1.2.2.2 Effects of Reforms on Education and Skills

Besides the mere effects of background characteristics on schooling and skill, education reforms are known to deploy policy instruments that impact education attained, and skills acquired. Interestingly, most policy instruments impact schooling and skill by influencing background characteristics (see Bjorklund and Salvanes, 2011). The study of Bjorklund and Salvanes (2011) gives insights into two broad categories of reforms that impact education and skills. Firstly, family-related reforms — e.g., preschool and maternity leave reforms (see Dustman and Schonberg, 2008); and school-related reforms—e.g., tracking and comprehensive school reforms (see Aakvik et al., 2010). This study exploits the latter (school-related reforms—the 1985 curriculum reform in Kenya) in a developing context.

As earlier highlighted limited related studies exist in the non-OECDs as most studies are in the OECD context, this study relates to the study of Aakvik et al. (2010) that explores the impact of mandatory education reform and background characteristics on educational attainment and earnings in Norway. Aakvik et al. (2010) found that the effects (on educational attainment) of background characteristics such as parental schooling and income were weaker after the reforms in Norway. Inspired by the work of Aakvik et al (2010), this study appraises the reforms in Kenya, examining the extent to which the background characteristics depend on the reform (and vice versa) in explaining educational attainment (years of schooling and skills in the form of reading proficiency). Hence, this study goes beyond the study of Aakvik et al. (2010) which considers the effects of background characteristics on educational attainment, exploiting the schooling reforms. As well as a consideration for the variation in background characteristics on schooling and reading proficiency, this study further unravels the effect of years of schooling (across credential categories) on reading proficiency, in an intergenerational framework. No single country in Kenya or sub-Sahara Africa has considered (jointly), the effects of background characteristics and schooling in an intergenerational framework. Besides the study of Aakvik et al (2010), this study also relates to the study of Meghir and Palme (2005) in Sweden; and Pekkarinen et al. (2009) in Finland. Both studies exploit school-related reform in Sweden and Finland. However, they have considered income as outcome, hence they have focussed on persistence or mobility in income across generations. However, this study exploits school-related reforms assessing persistence/mobility in schooling across generations; this study also examines equity in access to schooling and skills across the 'advantaged' and 'disadvantaged', ultimately, examining the efficiency of schooling—the extent to which schooling impacts skills in Kenya.

As earlier highlighted, this study is a unique blend of different strands of the economic literature and contributes to the broad literature on the economics of education, in developing contexts. Particularly, this study argues two (competing) approaches to raising learning outcomes—the equity and quality approaches—the former entails efforts to raise skills by considering background characteristics and raising (average) years of schooling. The latter involves raising skills by raising quality inputs in the schooling provision. Both approaches are deemed to have strong underpinning to background characteristics. However, whilst the latter is not empirically examined in this study, the former is assessed in two ways. Firstly, indirectly (research question 1); and secondly, directly (research question 2)—the former examines the effects of background characteristics on schooling and skill; and the latter involves examining 'direct' effects of education on skills, assessing the efficiency in schooling and the effects of background characteristics. Together both the direct and indirect approaches exploit variations in schooling, skill, and background characteristics attributable to the 1985 curriculum reform in Kenya. I now turn to discuss contributions, limitations, and the roadmap of this study.

2.1.3 Contributions

The primary contribution of this study is providing compelling empirical evidence that improves understanding of a more 'inclusive' and 'sustainable' strategy for raising learning outcomes[45] from investment in schooling, in developing contexts. Particularly, this entails support for equity in access to schooling (inclusion), minimising costs of investment in education (sustainability) over the current approach that entails increased quality inputs in

[45] More specifically, raising learning outcomes in this study entails developing reading proficiency from schooling.

schools that both inhibit inclusion and show high dependence on aid, hence, not very sustainable.

Amidst the substantial proportion of the disadvantaged in developing contexts, this study sets the scene for the effective use of policy instruments that can influence background characteristics to better raise average schooling across districts for a useful rise in skills, regardless of the quality of schooling. This study provides evidence for the argument against the 'current' approach that attempts to raise learning outcomes by raising the quality of schooling, through investment in school (quality) inputs. Whilst this latter (quality) approach may be useful in raising reading proficiency, in concordance with the literature, I argue that this is at best minimal, unsustainable, lacking support for equitable and substantial increase in schooling and skill, given the poverty in the region. Hence, in the quest to raise learning outcomes, this study provides evidence that accentuates background characteristics for schooling and reading proficiency. This strongly suggests that, in the place of mere policy instruments aimed at raising quality inputs, a consideration for policy instruments that influence background characteristics is crucial for raising the average (or aggregate) schooling and skills across districts in developing contexts. Whilst the extent to which the former (school inputs) impacts educational attainment is not empirically examined in this study, evidence from the literature strongly suggests most school inputs have no substantial effects, except for teacher quality, which notwithstanding is of relatively minimal effect on educational attainment – please, see the review of the literature above. Hence, owing to the peculiarity of background characteristics in sub-Sahara Africa, the former (approach for quality inputs) further raises the inequity in schooling (and skill) widening the gap in educational attainment (and skill) between the few advantaged and the many disadvantaged. The latter (consideration for backgrounds) not only improves the efficiency of resources but supports more equitable schooling and skill between the many disadvantaged and few advantaged. Ultimately, the latter approach has an overall effect in raising the aggregate schooling and skills across districts. Evidence presented in this work shows that aggregate schooling must reach a certain threshold (about twelve years of education) to result in favourable reading proficiency (and earnings—see Chapter 3) of individuals in urban Kenya. This supports the approach (as earlier argued) that raising skills through schooling is paramount for inclusive and sustainable growth, as an insufficient rise (below the threshold) in average schooling (adversely) impacts reading proficiency. This is particularly serious if school inputs do not guarantee a useful rise in educational attainment across districts.

In this study, both estimates of the effects of background characteristics on schooling and skills (research question 1); and estimates of the effect of schooling on skills (research question 2) exploit exogenous variations from the 1985 curriculum reform in Kenya. Particularly, in a Difference-in-Differences (DiD) analysis that relates to research question 1, the reform impacts schooling and skill through its influence on background characteristics which are shown to be crucial mechanisms for the effects of the reform on education and skills. For research question 2, this study also exploits variations in schooling attributable to the 1985 curriculum structural reform in Kenya, this forms the basis of the identification strategy that supports the causal inference on the effects of education on skill in the 2SLS-IV approach. Exploiting the 1985 curriculum reform helps to improve the internal validity of estimates in this study. The causal identification in this study helps to go beyond the useful descriptive evidence presented. This strongly supports (contributes to) policymaking.

Furthermore, besides the related work of Lucas and Mbiti (2012) that considers variability in test scores and access to primary schooling and the study of Ngware et al., (2006) which considers access in secondary schools in Kenya. This study does not only further examine the findings of Lucas and Mbiti (2012) and Ngware et al., (2006), this study goes beyond primary and secondary schooling and examines the effects of education on skills across all credential categories. Hence, this not only gives insights into the quality (efficiency) and equality in schooling in urban Kenya but also the use of a more comprehensive dataset with variables or factors—that include background characteristics of a broad swath of the country—that impact schooling and skills of the labour force (as opposed to the school-aged) whose credentials span all credential categories, improves the external validity of estimates, in this study. Hence, in this study, findings not only relate to policy on education but strongly impact the related policy on labour (or employment) and welfare.

Inspired by the study of Aakvik et al. (2010); Dee (2004); and Ganzach (2000), this study brings some methodological improvements to the table not only by examining crucial mechanisms through which schooling impacts skill or by assessing the sole effects of the reform and the effects of each of background characteristics on schooling and skill but a strong consideration of the variations in the effects of schooling on skill attributable to the joint or interaction effects of the reform and each of the background characteristics. Hence, in addition to assessing the mediation effect of each of the measures of background characteristics in the relationship between schooling and skill, this supports useful appraisal of the reforms in urban Kenya, by assessing the extent to which each of the measures of background characteristics depends on the reform. This further accentuates the mediation/moderation effects of other factors/variables that offset or synergise with the reform and the measures of background characteristics that impact schooling and skill. Particularly, the findings of these analyses add to the literature on intergenerational transmission of schooling and skill between parents and their wards, in addition to contributing to the literature that examines the efficiency of education, see Aakvik et al. (2010); and Lee and Wie (2017) for the latter.

The remainder of this chapter is organised as follows. Subsection 2.2 presents the Data and Methods subsection which also highlights the theoretical framework used to explore the research questions aiding the synthesis of evidence that supports the arguments (research questions) raised. Next, I present and discuss the Results subsection with some robustness checks in Subsection 2.3. Finally, in Subsection 2.4, I present the Summary of Findings and the Concluding Remarks including the useful policy implications of this study. Also highlighted in the final subsection are some limitations of this study and suggestions for future studies.

2.2 Methods and Data

Within the Methods and Data subsection, firstly, within the Methods unit, I present and discuss the theoretical framework, I then turn to the empirical framework subdivided into the empirical and identification strategies. Lastly, I present and discuss the data unit, including the descriptive evidence of key variables.

2.2.1 Theoretical Framework

The overarching objective of this study entails improving the understanding of an inclusive and sustainable approach to raising learning outcomes. This entails consideration for the effects of equity in access to schooling which involves examining the impacts of education on skill, accounting for background characteristics. As discussed in the Introduction the current approach to raising learning outcomes entails raising school quality inputs. This subsection lays out in functional forms (or mathematical representation), the relationship between variables based on the testable predictions of this study (see research questions 1 and 2). This also entails specifying the associated mechanisms and instruments through which causal inference (is drawn) between the dependent and the independent variables. Section 2.1 (subsection 2.1.2.1.2) presents reviews of the related literature on the theoretical framework adopted for this analysis. As earlier highlighted in the related literature (antecedents) subsection below, the theoretical framework of this study is inspired by the work of Bjorklund and Salvanes (2011) that emphasises the work of Becker and Tomes (1986) in support of utility-maximisation of the human capital theory. In this study, the approach followed is a slightly adapted version of the framework of Bjorklund and Salvanes (2011).

The two main relationships analysed in this study include the well-founded relationship between the educational attainment (of the offspring) and parental capital[46] via ‘intergenerational transmissions’ (see Equation 2.1); and the relationship between schooling and skill (see Equation 2.2). Hence, starting from the former, the next paragraph details the relationship between the educational attainment of the offspring (in their adulthood) as a function of ‘parental capital’ and related reforms that impact the education of offspring.

[46] In this study, ‘parental capital’ includes but not limited to parental schooling; possible abilities passed on genetically; and measures of parental wealth. Here, parental wealth is proxied by parental socioeconomic status at age 15 of the respondent which is deemed to be positively correlated to parental schooling.

Drawing from this relationship, the next subsection details the relationship between the schooling and skill of the offspring, highlighting the ‘parental capital’ and related reform as a crucial mechanism through which the education of the offspring is causal to their skill, in a developing context, specifically, in urban Kenya.

This study examines the extent to which reform results in variation between schooling and skill as a way of appraising the reform. Although the reform is deemed to be exogenous, I have included the reform indicator, X_r, in the relationships (Equations 2.1 and 2.2) with endogenous variables, this is done for ease of exposition, as the reform is deemed to result in (substantial) variations in the outcomes (schooling and skill).

Equation 2-1

$$S_r = f(S_m, S_f, A_r, f_p, X_r).$$

Here, the educational attainment, S_r of the offspring (respondent) which (in this study) is deemed to substantially explain the skills of the offspring (as a measure of their human capital), is a function of the following:

Here, a respondent’s choice to enrol and eventually complete a certain level (year) of education or credential category is most dependent on:

(1) the schooling of their parents (father and mother – S_m *and* S_f). Parental education (observed in this study) captures other unobserved factors such as parental risk, time preferences and parenting skills that are deemed to have (a causal) effect on the respondent’s choice to enrol and complete their current level of education.

(2) the ability (unobserved in this study) of the parents that is passed on to the respondent, A_r. This ability may be transferred genetically (nature) or environmentally (nurture) and potentially impact the cognitive skill of the respondents. As highlighted by the study of Bjorklund and Salvanes (2011), most schooling reforms may at best, have minimal impact on the effects of abilities transferred genetically. An example is that, whilst health reforms may make it possible to diagnose and address possible adverse effects of genetically transferred traits (from parents to the offspring). This may be a useful way to improve educational attainment, however, schooling reforms alone may be inadequate and at best be complementary to health reforms in such cases.

(3) socioeconomic status (f_p) (observed in this study). This is a measure of the wealth of the respondents' parents (at age 15), this gives insights into the well-being of the offspring, in their formative years. Although this is deemed to have a strong correlation to parental education, however not in all cases, as parents may have received their wealth from their parents and not necessarily due to their schooling. However, parental schooling and wealth are of interest for this study in sub-Sahara Africa, as earlier highlighted (see antecedents in section 2.1), as evidence suggests schooling is much dependent on parental schooling and wealth, as public provision of education is limited on several fronts.

(4) the reform indicator, X_r, a measure of all public policies that impact the public provision of schooling, including all school inputs that may explain educational attainment. In this study, the reform is deemed to interact with all the other observed explanatory variables discussed, and these interaction terms are deemed additional mechanisms or factors through which parental capital and the reform impact schooling. Hence, put together (as in Equation 2.1), the relationship shows that a respondent's educational attainment is explained by their parental capital, the reforms, and interaction terms of parental capital and reforms.

Modelling the Mechanisms through which Schooling Impacts Skill.

Inspired by the work of Bjorklund and Salvanes (2011), the following (Equation 2.2) depicts the relationship between reading proficiency as a measure of (cognitive) skill (hence a measure of the human capital of the respondents) (H_r); and the 'parental capital (H_p)', the unobserved abilities of the respondents (A_r); and the reform indicator (X_r) as earlier described in Equation 2.1. Also, the parental capital (H_p) entails parental schooling (mother, S_m; and father, S_f) and wealth (f_p), as earlier described in Equation 2.1.

Equation 2-2

With this in place $H_r = f(H_p, A_r, S_r, X_r)$.

In this relationship, the following are assumed:

$\frac{\partial^2 f}{\partial S_r \partial H_p} \geq 0$; $\frac{\partial^2 f}{\partial S_r \partial A_r} \geq 0$; and $\frac{\partial^2 f}{\partial S_r \partial X_r} \geq 0$.

Here, a positive relationship exists across parental capital (H_p), abilities (A_r), and the reform (X_r). Useful schooling reforms over time raise the parental capital $(S_m, S_f\ and\ f_p)$ which are correlated. These, in turn, raise abilities. These variables (including the interaction terms of background characteristics and reforms) explain access to schooling and educational attainment that impacts skills. These are factors and mechanisms through which changes in levels of education explain changes in skill levels. However, in this study, as abilities are unobserved, this at best validates the following:

$$\frac{\partial^2 f}{\partial S_r \partial H_p} \geq 0; \frac{\partial^2 f}{\partial S_r \partial X_r} \geq 0; \text{ and } \frac{\partial^2 f}{\partial H_p \partial X_r} \geq 0$$

These assumptions suggest the mechanisms (or factors) through which schooling impacts skill is via reforms, parental capital $(S_m, S_f\ and\ f_p)$, and interaction terms of the reform and each of the measures of parental capital. Both (1) having wealthy or educated parents; or the reforms in place and (2) having educated parents with reforms in place, make factors or conditions ((1) and (2)) that raise skills by raising the schooling of the respondents. Hence, not only do mere parental schooling, mere parental wealth or mere reform impact the schooling which impacts the skill of individuals as in (1), but (2) suggests a possibility of reforms in place to impact schooling or wealth of parents that in turn impact the schooling and hence, the skill of the offspring. (1) makes the unique effects of each of the variables and (2) makes the interaction effects of the variables as factors that impact the relationship between the schooling and skill of the respondents.

Both equations 2.1 and 2.2 make a recursive system that gives useful theoretical underpinning to the analysis in this study that exploits variations in schooling, parental capital, and skill in the outlined framework. I now turn to the Data subsection.

2.2.2 Data

This data subsection provides a brief introduction to the STEP survey used in this study and a description, descriptive evidence, specifications, and some descriptive analyses of the key variables of this study.

2.2.2.1 Introduction to STEP Surveys for urban Kenya

This study uses the World Bank's STEP[47] Skill Measurement surveys for Kenya. The Household Survey (HS) was fielded from 01/08/13 to 30/11/2013 as part of the second wave of the STEP Skills Measurement Programme. The STEP is the first initiative to measure internationally comparable skills, in low- and mid-income (non-OECD) countries. It elicited data from adults between the ages of 15 and 64 in urban Kenya. Modules in the STEP HS comprise demographic and dwelling characteristics; data on education, training, and reading literacy test assessment; data on employment, health, and job skill requirements; personality behaviour and preferences; and finally, data on language and family background characteristics. The STEP data aid useful understanding of education attainment and skills as measures of human capital in the non-OECDs; and how human capital impacts productivity or earnings in labour markets (see Nogales and Krishnakumar, 2020; Valerio et al., 2016). The datasets include a wide range of measures of human capital such as educational attainment, skill (including personality traits), and background characteristics (including socioeconomic status and parental education) that can test hypotheses that relate to the objective of this study.

Although this study focuses on Kenya, there are about seventeen (17)[48] STEP participating countries. The STEP HS for urban Kenya has 1,196 variables with a survey sample size of 3,894 respondents (observations). The STEP household survey has lineage to the OECD's PIAAC – Programme for the International Assessment of Adult Competencies available in twenty-four high-income countries (OECDs). The PIAAC surveys include about 166,000 respondents, aged 16-65 years, across the participating countries. Hence, the STEP dataset makes it possible to undertake comparable research on education (schooling) and skill, in the non-OECDs as those that have used the PIAAC. The main survey sample of the STEP HS is defined by the World Bank sampling (and weighting) methodology (see the detailed

[47] STEP – Skills Toward Employment and Productivity.

[48] Albania, Armenia, Bolivia, Bosnia & Herzegovina, Colombia, Georgia, Ghana, Kenya, Kosovo, Lao PDR, Macedonia, Serbia, Sri Lanka, Ukraine, Vietnam, and Yunnan Province in China.

Data Section in the Appendix Chapter). The focus is on the main survey sample as the analytical sample of this study, also considered are subsamples that test heterogeneity across background characteristics. The survey is restricted to urban Kenya. Hence, conclusions from findings using the STEP dataset refer to ‘urban Kenya’ and not ‘Kenya’.

As earlier mentioned, the STEP survey collected data for the working-age (15–64, inclusive) population. Unlike many other studies, the sample for this study is not limited to a certain age range deemed more economically active, as part of the objective is to examine the (full) growth impacts of education and skill. The government of Kenya has a statutory age for entry and retirement, especially for those in public service employment. However, this age restriction is seldom followed by other sectors. It is typical to commence full-time work before reaching the legal working age; and leave active employment at much later ages. Hence, rather than alter the age (lower and upper boundaries), I do not raise the entry age or lower the exit age but rather use the entire age limit (15-64 inclusive) of the labour force in the World Bank STEP dataset for Kenya. The requirements (based on the objective of this study) of examining the impact of background characteristics on education and skill make the use of sub-samples of the analytical sample salient in this analysis.

2.2.2.2 Variables – Descriptive Evidence and Specifications

This study aims to provide evidence in support of increased access to schooling with an emphasis on the effects of background characteristics as key explanatory variables for skill proliferation. Hence, this study considers measures of schooling, parental education, and socioeconomic status, as key variables that explain skills. However, the very nature of skill is such that several other variables including neighbourhood and peer characteristics as earlier emphasised (see literature reviews) and factors such as age, training, and measures of work experience are other measures that may strongly impact skill or reading proficiency. In this study, in addition to factors such as neighbourhood or peer effects[49] as discussed in subsequent subsections and chapters of this work also accounted for are, the age, training and work experience of respondents. Examining the effects of these variables is known to impact skills and improve estimates of the effects of key variables of interest in this study. Hence, the measures of background characteristics and schooling are deemed to be more precisely estimated by accounting for the age, training, and work experience of respondents in this study. However, some related studies account for the potential experience, which may

[49] These variables include innovations from the stratum variables described in Table 2-1, they include measures of average schooling and skill across districts.

be appropriate in the OECDs where entry and retirement from the labour market are mainly fixed by school-leaving age and the related statutes. In this study for urban Kenya, instead of potential experience, the actual number of months of experience in the current role (tenure) and its quadratic term (tenure_squared) are controlled for. This is crucial in research in the non-OECDs where the age of entry to (and exit from) the labour market varies significantly, with the prevalence of child labour (and related statutes that are seldom complied with) in the non-OECDs, see Manda et al. (2003) for some evidence in Kenya. In addition to this, the choice of tenure (and tenure_squared) instead of potential experience also relates to issues with female employment where entry to (and exit from) the labour market varies and relates to religion/tradition or childbearing, amidst the rise in female education across the world (Klasen, 2019). I now present a brief description and some descriptive evidence of key variables of this study, after which, I present and discuss more detailed variable specifications and further descriptive evidence of some key variables as earlier highlighted.

Key Variable Description and Descriptive Evidence

Table 2-1 Key Variables: Brief Description and Some Descriptive Evidence

Variable	Brief Description	Obs	Mean	Std. dev.
years_educ_act	Continuous, actual number of years of schooling	3,283	10.604	4.222
apvlit_c	Continuous, cognitive skills – reading proficiency.	3,301	178.808	85.347
apvlit_d	Categorical, cognitive skills – reading proficiency.	3,261	1.123	0.973
ISCED	Categorical, highest qualification attained (credentials)	3,285	2.426	1.454
training	Continuous, participated in training in the last year	3,301	0.117	0.321
age	Continuous, age of the respondent.	3,301	29.597	9.919
age2	Continuous, quadratic term of age.	3,301	974.303	722.763
gender	Dummy, an indicator of female gender.	3,301	0.527	0.499
father_educ.	Categorical, an indicator of the father’s level of education.	3,301	1.557	1.037
mother_educ.	Categorical, an indicator of mother’s level of education	3,301	1.250	0.967
SES	Categorical, indicator of SES at age 15	3,289	1.876	0.582
p1985_	Dummy, the reform indicator (instrument 1, coding 2)	3,174	0.897	0.304
m1a_q05m	Categorical, indicates the month of birth (see QoB).	3,295	4.179	12.600
m1a_q05y	Continuous, indicates the year of birth	3,298	1983.187	9.896
tenure	Continuous, number of months of experience	2,063	54.722	61.105
tenure_squared	Continuous, quadratic term of tenure.	2,063	6726.497	16852.44
school _location	Categorical, includes, another_city, same_city (reference) and foreign_city	3,162	1.621	0.514
school_type	Categorical, includes public (reference), private, and other (inc. homeschooling)	3,171	1.316	0.680

Note: The evidence is from the Main Analytical Sample of the STEP Household Survey for urban Kenya. Please see further details and descriptive evidence of each of these variables in the subsequent paragraphs and the Data Sections of the Appendices. Please see the detailed specifications of the stratum variables and the measures of aggregate schooling and skills in Chapter 4.

Variable Specification and Further Descriptive Evidence:

2.2.2.2.1 Cognitive Skills as ‘Skill’ or Reading Proficiency

In this study, the term cognitive skill ‘skill’ has been used extensively as a measure of human capital. However, the study of Borghans et al. (2001) emphasised the need to be clear with the term as its use is more susceptible to misinterpretations these days than ever, given the emergence of new understanding and ideas of varying forms of ‘skill’ over time. The use of outcomes of literacy, numeracy, and problem-solving tests make useful measures of human capital, over time, these measures have gained consensus as measures of cognitive skill in Labour and Education Economics and the wider Social Sciences (see Hanushek et al. (2015; and 2013)). Within the STEP dataset are the indirect and direct measures of cognitive skills. The indirect measures are self-reported and subject to criticism. However, assessing adult reading proficiency by ‘direct measurement’ of cognitive skills is deemed a more credible measure of cognitive skills. The assessment is administered by the STEP team and designed by ETS – Educational Testing Service. Unlike the questionnaires for the self-reported (indirect) measures of cognitive skills, the assessment comprises four sections, and these include Comprehension Passage; Sentence Processing; Vocabulary; and a Core Section (literacy assessment) that elicit skills requiring the respondents to demonstrate useful levels of interpretation, identification, and planning. Typical of large-scale assessments which have a large number of items, due to limitations in time, respondents only attempt a subset of the items in the assessments, this may result in measurement error in the assessments of the reading proficiency of the respondents. The Plausible Values (PVs as several imputations or scores) with a scale ranging from 0 to 500 and using the Item Response Theory. The PVs minimise measurement errors and provide a more reliable assessment of the reading proficiency of respondents. The following Table presents the summary statistics of each of the PVs for reading proficiency in urban Kenya. In the STEP data, 10 PVs represent the measure of the reading proficiency of each of the respondents.

There are three main approaches in specifying the PVs of the 'skill' variable. These include: (1) the use of a single PV out of all the PVs, this may be the first or a randomly selected PV among the PVs; (2) an aggregate or average term from all the PVs; (3) an average of the estimates obtained using each of the PVs (or a few, such as five PVs that are randomly selected) individually (see Laukaityte and Wiberg, 2017). The study of Laukaityte and Wiberg (2017) concluded that (1) leads to a bias as the parameter estimate from the use of a selected PV varies depending on which PV is selected. Using the multilevel modeling (MLM) approach and the TIMSS[50] 2011 dataset Laukaityte and Wiberg (2017) also warned against the use of the average of PVs (2), which they concluded as giving similar results as the use of the average from each of the PVs (3) – as doing so may result in less precise estimates of the population parameter – evident from the differences in the standard errors and within-school variances. Ultimately, they warned against the use of a single (1) and the use of the average of PVs (2) owing to issues of precision in estimates as (2) in particular, results in underestimation of the standard errors. On the use of (3) below, the work of Laukaityte and Wiberg (2017) further suggests, increasing the number of PVs used from the usual practice of the use of five PVs can increase the precision of estimates of population parameters in some cases. However, they offered no clear suggestions as to implementing this. Ultimately, the work of Laukaityte and Wiberg (2017) regards all three approaches outlined below as resulting in imprecise estimates of the population parameter. However, the relatively recent empirical work of Bibby (2020) suggests that the use of five PVs (of (3) below) remain adequate in the estimation of population parameter. In addition to this, it is important to note that the study of Laukaityte and Wiberg (2017) has used the TIMSS (a school-based study/dataset) in arriving at the weaknesses of (2)—that entails the use of the average of all imputations (or PVs)—and (3) below, which they claim gave implausible

[50] Trends in International Mathematics and Science Study (TIMSS).

estimates owing to differences in standard errors and within-school variances. Furthermore, it is important to note that the STEP data used in this study involves an assessment of adult reading proficiency, but the TIMMS is an assessment of the cognitive skills of the 'school-aged'. Hence, relative to the STEP data, clustering in the TIMMS may have implications for the robustness of estimates. This may explain possible differences in the standard errors and within-school variances evident in working with TIMMS as Laukaityte and Wiberg (2017) suggest. Moreover, the more recent empirical analysis of Aparicio et al. (2022) suggests, that regardless of the settings—school-based or otherwise—all three of the methodologies for handling PVs did not result in any material differences in estimates of the population parameters. For the variable specification in this analysis, I have used the average of all PVs (2) not (1) or (3) below. Although (1) appears to be simple. However, the argument of Laukaityte and Wiberg (2017) which concluded that the use of (1) leads to a bias, is appealing enough. Although (3) seems to have gained reasonable support from the work of Bibby (2020), Laukaityte and Wiberg (2017) suggest increasing the number of PVs used may raise the accuracy of estimates. Interestingly, the empirical work of Araki (2020) used (3) implementing all the 10 PVs of PIAAC for Japan and Korea, affirming that although it was a tedious process, it was worth it, as he attested to the approach.

However, one would expect possible errors from taking the averages of multiple outcomes (as in (3)). Moreover, (2) seems to provide a more simplistic and accurate approach (Aparicio et al., 2022) to specifying the 'skill' variable in simple (or non-multilevel) models used in this study, unlike the multilevel models used by Araki (2020) and referred to by Laukaityte and Wiberg (2017). Hence, in specifying the skill variable used in this study following (2), I take the average of the 10 PVs (pvlit1 to pvlit10) to give, apvlit_c, typical of most analyses involving measures of skills, I standardise apvlit_c (average of all PVs) to give, zapvlit_c used in all econometric specification, unless stated otherwise. However, the apvlit_c (non-standardised) has been used for ease of exposition in this data subsection. The use of **c** as seen in the aggregated variables tells the variable is continuous, rather than categorical. In the subsequent analysis, the **d** for the PVs indicates a dummy or categorical variable.

For the following charts, the y–axis shows densities; and the x-axis shows apvlit_c, with a range of 0-500 across distributions.

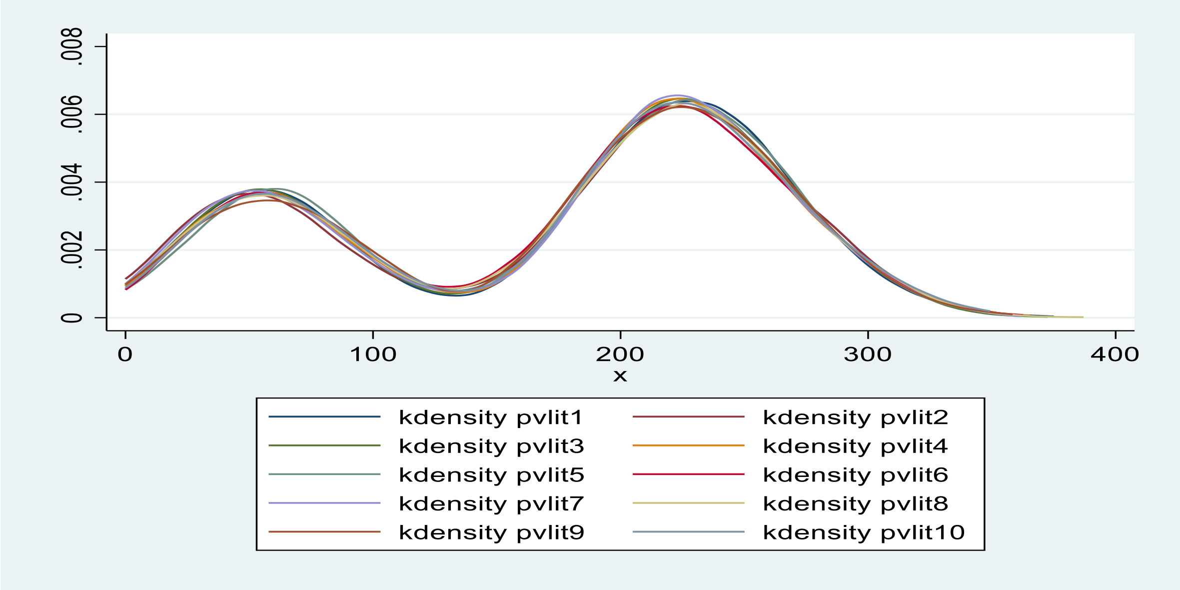


Figure 2-4 Kernel Density Plots, Distribution of Reading Proficiency.
Source: STEP Skills Measurement Program of the World Bank.

Table 2-2 presents some descriptive evidence of each of the 10 PVs and the average for urban Kenya. Table 2-4 shows that the distribution of all PVs. The mean of the PVs appears to vary from the median PV of the distribution in urban Kenya. This suggests an asymmetric distribution across all samples/subsamples.

Table 2-2 Summary Statistics of Plausible Values

Plausible Values	Obs	Mean	Std. dev.	Min	Max
pvlit1	3,301	178.639	87.448	0	375.028
pvlit2	3,301	178.02	88.430	0	376.012
pvlit3	3,301	178.711	87.153	0	372.534
pvlit4	3,301	178.843	87.841	0	365.255
pvlit5	3,301	178.995	86.446	0	368.336
pvlit6	3,301	179.425	86.162	0	366.309
pvlit7	3,301	177.835	87.328	0	361.766
pvlit8	3,301	178.372	87.559	0	387.041
pvlit9	3,301	179.086	87.009	0	358.395
pvlit10	3,301	180.157	86.832	0	353.074
apvlit_c	3,301	178.808	85.347	14.169	350.919

Notes: Table of summary statistics for Plausible Values – reading proficiency of respondents – main Analytical Sample, STEP HS for Kenya

In urban Kenya, the mean reading proficiency (apvlit_c) is 178.8083 out of 500. Next to Kenya is Ghana, with an average of 135/500 in reading proficiency, this makes Ghana the country with the least average reading proficiency, among the STEP participating countries. However, using the comparable PIAAC for the OECDs, the average reading proficiency for Japan is about 297/500.

To further analyse skill, using the continuous measure of skill, apvlit_c as described, I derive (or code) a categorical variable (measure) of skill (apvlit_d) by assigning 0 to 4 to each of the defined reading proficiency levels (based on the Plausible Values, PVs as earlier described). The following defined categorisation of reading proficiency is recognised by international surveys that deploy the Plausible Value methodology of measuring skill. These international surveys include (but are not limited to) the PIAAC, STEP, TIMSS, and PISA as earlier described, including the PIRL—Progress in International Reading Literacy Study.

The categories (STEP categorisation) are defined thus:

Level 0 (Reference): 0—175; Level 1 (Basic): 176—225; Level 2: 226—275 (low-level inferences); Level 3 (some complex inferences): 276—325; Level 4 (complex inferences): 326—375; Level 5 (high-level inferences or constructing synthesis): 376—500. However, no respondents attained Level 5 reading proficiency in the STEP HS for urban Kenya. Hence, in coding the categorical measure of reading proficiency apvlit_d, I assign 0—4 instead of the 0—5 categories as thus: Level 0 (Reference):0—175; Level 1 (Basic):176—225; Level 2:226—275 (low-level inferences); Level 3 (some inferences):276—325; Level 4 (complex inferences): 326—500. I then created the five dummy variables used in this analysis.

2.2.2.2.2 Socioeconomic Status

Table 2-3 Categories of Socioeconomic Status

SES	Socioeconomic status	Freq (Percentage)
1	Low SES (ses_1)	786 (23.90)
2	Mid SES (ses_2)	2,124 (64.58)
3	High SES (ses_3)	379 (11.52)
Total		3,289 (100)

Note. Source: Main Analytical Sample of The World Bank’s STEP Data for Kenya.

The socioeconomic status variables (see Table 2-1) are specified from a self-reported 10-scale (with 10 for richest and 1 for poorest) variable on the respondent’s perception of the economic status (wealth of family) when the respondents were 15 years of age. From this, a 3-level categorical socioeconomic status variable is specified with 1 as the lowest; 2 as mid-level, and 3 as the high socioeconomic status of the respondents. From this, I create two dummy variables (ses_3, as high SES; and ses_1, as low SES). Based on this specification of the SES in urban Kenya, under 12% of urban Kenya had high SES, and double (24%) had low SES in urban Kenya. About 65% had mid-SES growing up. This category serves as the reference category and doubles the sum of those with high and low SES.

2.2.2.2.3 Education and Structural Reform

Measures of Educational Attainment

The STEP provides data on the educational attainment of both parents individually, for mother and father, detailing the highest qualification or level of schooling attained. The STEP also includes extensive data on the education attainment of all respondents – this includes, data on the expected years of schooling of the respondents (years_educ) which is obtained from the data on the highest qualification attained, also provided by the STEP is data on actual number of years of schooling of the respondents (years_educ_act) – relative to the former, the latter is useful in this analysis as it accounts for grade repetition, which is peculiar to educational systems in sub-Sahara Africa (inclusive of Kenya), where students are required to repeat years of schooling when they performed below certain thresholds.

Before the 1985 reform primary education (should) normally commence at age 6, starting from Grade 1 to Grade 7 when pupils sit the Certificate of Primary Education (CPE) exam. At about age 13, most pupils commence their lower-secondary schooling that spans Forms 1 - 4, at the end of Form 4, pupils sit the Certificate of Secondary Education (CSE) examination and progress to upper-secondary schooling at age ≈17. The upper-secondary schooling spans Forms 5 – 6. By the end of which students sit University/College Entrance examinations. Typically, university education spanned three (first, second, and third) years, Typically, this should commence at age 19 running through to 21. Post-1985 Reform saw some changes as thus: Primary education spans Standards 1-8, typically, commencing at age 6 and terminating with the Kenya Certificate of Primary Education (KCPE) at age 13. Hence, at age 14, pupils commence secondary schooling that spans Forms 1-4 terminating with the Kenya Certificate of Secondary Education (KCSE). The university education typically spans four years, commencing from age 18 and running through to 21. Hence, there are no material changes (in the duration) for those who complete tertiary schooling, however, for those who complete primary and secondary schooling (and the respective credentials) there are material changes in their years of education. I have highlighted the expected ages of entry and exit across credential categories, in Kenya. However, there are significant variations in the ages of commencement and completion. This is attributable to the high rates of grades/form repetitions (Somerset, 2007 and 2009) and according to Somerset (2007), as of 1978, more than half the pupils enrolled in grade 1 are not six (6) years of age which is the expected age at first grade. This suggests material issues of access to schooling (Chicoine, 2009; Somerset, 2007). Hence, some peculiarities of education in Kenya (and the sub-Sahara Africa) such as grade repetitions; and the variance in *age at first grade* make determining birth cohorts a tough task.

Table 2-4 Difference between Expected and Actual Years of Schooling

Variables	Brief Description	Obs	Mean	Std.Dev	Min	Max
years_educ_act	Continuous, actual years of schooling	3,875	10.32	4.27	0	22
years_educ	Continuous, expected years of schooling	3,868	9.25	4.73	0	22

Source: Author's elaboration of the STEP data for Kenya.

Table 2-4 shows that the difference between actual and expected years of schooling is about 1.07 years, suggesting that, on average, people spend an extra year of schooling above what is expected to attain their current credentials/qualifications. The average number of years of schooling in urban Kenya is 10.32 years.

Table 2-5 Age at First Grade and Age of Respondents

Variables	Brief Description	Obs	Mean	Std. D	Min	Max
age_start	Discrete, the actual age of respondents at first grade.	3,733	6.69	1.05	3	18
Age	Continuous, actual age of respondents.	3,894	29.53	9.93	15	64

Source: Author's elaboration of the STEP data for Kenya.

Although pupils are required to commence schooling at age six (6) however, Table 2-5 indicates a substantial proportion of the pupils commence at much later ages. Interestingly, a few commence earlier than 6 years of age. The variance between the expected (6 years) and actual age at first grade (substantially, higher than 6 years, up to 18 years) and grade/form repetitions not only evidence issues of access to schooling as earlier highlighted but these constrain understanding of useful birth cohorts of the respondents. Hence, these make it tough to know those impacted by the 1985 reform resulting in complications in specifying the reform indicator used extensively in this study.

Besides the continuous measures of schooling for the respondents, the STEP Household Survey also provided categorical measures of schooling for respondents. However, the STEP only provides a categorical variable that captures parental schooling (of respondents). The educational attainment of parents and respondents is classified using the ISCED[51]. The ISCED is increasingly important in research as it aids the comparability of credentials across world regions, using an internationally comparable credential classification system. In the STEP data set, the isced, mother_educ, and father_educ variables are measures of categorical schooling for respondents, their mothers, and fathers, respectively. These variables are used extensively in this chapter. As earlier highlighted, the STEP dataset uses the ISCED 1997 to achieve comparability of educational attainment across the STEP participating countries and similar adult literacy surveys like the OECD's PIAAC (see Appendix for detailed data description). Generally, the ISCED 1997 includes the following categories: ISCED0—pre-primary education; ISCED1—primary education; ISCED2—lower-secondary education; ISCED3—secondary and some post-secondary education; ISCED4—post-secondary (advanced/technical) but non-tertiary; ISCED5—first-stage tertiary (undergraduate/postgraduate taught) education; and ISCED6—second stage tertiary (advanced research) education.

Table 2-6 Categories of Educational Attainment for Respondents, Father and Mother

	isced	**mother_educ**	**father_educ**
isced01 (no qualification attained from 0 – few years of schooling)	350 (10.65)	845 (25.60)	696 (21.08)
isced1 (completion of primary schooling)	694 (21.13)	1,177 (35.66)	742 (22.48)
isced2 (completion of lower-secondary schooling)	436 (13.27)		
isced2&3 (completion of lower- and upper-secondary schooling)		889 (26.93)	1,191 (36.08)
isced34A (completion of upper-secondary schooling)	1,157 (35.22)		
isced4B (post-secondary advanced but non-tertiary schooling)	309 (9.41)		
isced4 & Higher (technical/advanced but non-tertiary)		390 (11.81)	672 (20.36)
isced56 (tertiary education only)	339 (10.32)		
Total	3, 285 (100)	3,301 (100)	3,301 (100)

Note: Table 2-6 is the Author's elaboration of the STEP data for Kenya. Table 2-6 shows some further descriptive evidence of key categorical schooling variables for the respondents (isced), their mother (mother_educ); and father (father_educ). Figures without brackets are the frequencies and those with brackets are the respective percentages. Hence, each column adds up to 100%.

[51] ISCED – the International Standard Classification of Education. Also used in the OECD's PIAAC, the ISCED is the reference international classification for organising education programmes and related qualifications by levels and fields, providing a basis to compare qualifications across the world. Specifically, the STEP for Kenya uses the ISCED-1997.

Table 2.6 suggests fathers are more educated than mothers, just as more male respondents have tertiary education, relative to the female respondents. From the *isced*, *mother_educ*, and *father_educ* variables in the STEP HS (see Table 2.1) detailed below, I code the related dummy variables used extensively in this study. The evidence suggests over 20% of fathers have post-secondary qualifications and 12% of mothers are in the same credential category. Regardless of gender, under 11% of the respondents completed tertiary education. However, from the trend for those with no credentials (this may include respondents and parents with few years of schooling), evidence suggests this is declining as over 20% of the parents regardless of gender, have no credentials. However, for respondents (later generation), only about 10% have no credentials.

Specification of the Reform Indicator—the reform dummy, p1985_ and the Quarter of Birth; and Balance Test using the reform dummy.

Inspired by the study of Patrinos et. al. (2021) that estimated the reform-affected returns to education in Turkey, I specify the 1985 reform indicator using a similar approach and following a review[52] of the implementation of the structural reform in Kenya, at the beginning of 1985. The actual number of years of schooling, instead of the expected years of education accounts for issues of grade repetition as earlier highlighted, the variances in age at first grade further complicate the determination of the birth cohorts impacted by the reform. However, using the 'age_start' and 'years_educ_act' variables (see Tables 2.4 and 2.5) offers insights from which a useful reform indicator is specified. The study of Chicoine (2012) uses five rounds (1989, 1993, 1998, 2003, and 2008) of the Demographic and Health Surveys (DHS) conducted in Kenya to specify a policy dummy using the 1985 curriculum reform in Kenya. To overcome complications presented by grade repetition and the variability in age at first grade that should determine the birth cohorts impacted by the reform, Chicoine (2012) uses several measures in determining cohorts influenced by the 1985 curriculum reform in Kenya. Measures used include the pre-reform data for those enrolled in Grade 1 including data that capture age in Grade 1; data on enrolment and repetition at every grade of primary school and; a transition cohort ($1965Q1 \leq$birth cohorts$\leq 1971Q4$) from the last year of primary to secondary school. However, in this study, in specifying the reform dummy, p1985, rather than following the approaches of Chicoine (2012), the STEP HS provides the 'age_start' variable, indicating the actual age of respondents at the start of their first grade; and the 'years_educ_act' indicating actual number

[52] See the Introduction where the 1985 curriculum reform implementation is discussed.

of years of schooling for the respondents. These variables make it possible to identify the birth cohorts impacted by the reform, accounting for the variability in age at first grade and possible repetitions by respondents, this makes it possible to innovate, specifying a useful indicator of the 1985 reform (please refer to the 1985 curriculum reform implementation in the Introduction and the Appendix) as thus:

Pupils not enrolled in secondary school by Quarter 4 of 1984 had an extra year added to their primary schooling from Quarter 1 of 1985, hence, instead of the usual seven (7) years of primary education, they had eight (8) years of primary schooling. Assuming no repetitions and no variability in the age of commencement (where pupils commence schooling at six (6) years of age); and have seven (7) years of primary education as in the old regime. Hence, at a maximum age of 13 (at 1984Q4)[53] working with the year of birth variable (m1a_q05y)[54] means that a pupil would have to have been born by the end of 1971Q4, for the pupil not to have eight (8) years of primary schooling (hence, not impacted by the new reform). This means that to be affected by the reform, a pupil would have to have been born by the first quarter of 1972 (Q1) and thirteen (13) years from 1972Q1 would take the pupil to when the reform was operationalised (January, 1985Q1). Hence, this makes 1972Q1 the reference year of birth for the cohort impacted by the reform. Therefore, the reform indicator is specified based on Equation 2.3 as follows:

Equation 2-3

1972Q1 = 1985Q1–13; Or 1972Q1 + 13 = 1985Q1

With no complications – hence, without issues of grade (or school year) repetitions; and variability in age at first grade – the following reform indicator (_p1985) makes a useful reform dummy as thus:

$$\text{\_p1985} = \begin{cases} 1 \; if \text{ m1}a\_q\text{05y} \geq 1972Q1 \\ 0 \; if \text{ m1a\_q05y} \leq 1971Q4 \end{cases}$$

[53] The age 13 is made of six (6) years of age after birth + seven (7) years of primary schooling.

[54] The STEP dataset contains the year of birth variable, m1a_q05y (year of birth of respondents) where respondents born within the range of years (1948 – 1998) as it was fielded in 2013 and only includes respondents aged (15-64, inclusive) years.

Table 2-7 Descriptive Evidence of Reform Indicator (coding 1), _p1985.

Variables	*Brief Description*	*Obs*
_p1985	Dummy, reform indicator.	3,298
0	Control	415 (12.58)
1	Treated	2,883 (87.42)

Source: Author's elaboration of the STEP data for Kenya.

However, with the reality of the substantial grade repetition (or dropout and return during schooling) and variation in age at first grade in Kenya (please, see the differences expected and actual number of years of education and; the descriptive evidence of age_start in tables 2.4 and 2.5), the coding 1(_p1985) will not account for complications such as grade repetition and the variability in age at first grade. However, given these complications, I made some innovations, modifying the year of birth variable, m1a_q05y, by adding, the actual age at first grade, age_start to the year of birth, then adding 7 as the number of years of primary schooling pre-1985 reform. This creates a new variable, _ref from which I create a new reform indicator, p1985_ as thus:

Equation 2-4

_ref = m1a_q05y + age at first grade + 7.

Hence, in place of Equation 2.3, and a cut-off at 1972Q1, Equation 2.4 with the actual reform date of 1985Q1 are used in this instance.

Therefore,

$$\text{p1985\_} = \begin{cases} 1\ if\ \text{\_ref} \geq 1985Q1 \\ 0\ if\ \text{\_ref} \leq 1984Q4 \end{cases}$$

The 'seven years' of primary schooling in the old regime is mainly the case for most respondents (deemed to have been impacted by the reform), in their years as pupils in urban Kenya. However, assuming most respondents impacted by the reform only had seven years of schooling may result in not fully accounting for possible grade skipping or repetitions and intermittent dropouts[55] in the primary schooling of the respondents. This may result in substantial errors in assignments to treatment and control. However, using actual years of schooling (instead of the expected number of years of schooling) accounts for repetitions, grade-skipping, and possible intermittent dropouts, which otherwise may not be accounted for with expected years of schooling. However, this approach (coding 2) does not fully eliminate errors in assignments to treatment and control, particularly, where continuity across levels of schooling is not a probable assumption in sub-Sahara Africa where several factors (including financial factors) materially impact time spent in school. Notwithstanding, relative to coding 1, this approach (coding 2) substantially mitigates defects in assignment to treatment. Somerset (2007) finds that, in Kenya, grade repetitions were greatest in Grade 7, with repetition across Grades 1-7 for 1974-1978 averaging at about 5.7% in Kenya. Although the WB STEP collected the data ***interrupt***, indicating only 4 out of 3,894 respondents did not interrupt their schooling for at least an academic year, however, only 7% of respondents provided information on their exact number of years of interruption. This inhibits further innovation in coding 2 that could further correct the defects of coding 1. Coding 2 (p1985_) is the preferred specification of the reform indicator used in the study[56]. The descriptive evidence of p1985_ (see Table 2.8) suggests that the reforms impacted about 90% of the respondents in the analytical sample.

Table 2-8 Descriptive Evidence of the Reform Indicator (coding 2), p1985_

Variables	*Brief Description*	*Obs*
*P1985_*	Dummy, reform indicator.	3,174
0	Control	326 (10.27)
1	Treatment	2,848 (89.73)

Source: Author's elaboration of the STEP data for Kenya. The figures in brackets are percentages and those without brackets are frequencies of the treated and control groups.

Besides the use of coding 2 as an instrument for schooling and skills, the month of birth variable, m1a_q05m (see Table 2.1 for a brief description), m1a_q05m is used to create (code) the Quarter of Birth (qob) used as the second instrument of schooling. The qob is coded by re-assigning each of the months into quarters by doing the following: 1/3=1;4/6=2;

[55] Here, intermittent dropouts (and returns) refer to situations where pupils leave school for a while and return at future dates. This may result in wrongly classifying a student into the control (by merely adding 7 years to their age at first grade) instead of the treatment category. Another is cases where the 'very good' or the 'gifted' spend less time in primary (or other categories of) school by skipping grades. These are the inherent defects of the approach considered.

[56] This is subject to further testing within the identification strategy subsection. Please, see the empirical framework.

7/9=3;10/12=4. Where 1 is month 1 and 12 is month 12, deleting unassigned observations (or observations) with a missing month of birth (such as those assigned to 66). The study uses up to three instruments in the identification strategy pursued using the quasi-experimental approaches that entail using the instruments to exploit (exogenous) variation in schooling and skill. These instruments include the reform indicator (coding 2) and the interaction of the reform indicator and dummies for quarters of birth as the second instrument. Inspired by the studies of Acemoglu and Angrist (1999), the Instrumental Variables strategy deployed uses the interaction of the quarter of birth variable derived from the month of birth variable (m1a_q05m); and the reform dummy. However, the Difference-in-Differences approach uses the reform indicator solely.

Table 2-9 Descriptive Evidence, Quarter of Birth (QoB)

Quarter of Birth	Frequency	Percentage
Q1	784	24.51
Q2	945	29.54
Q3	785	24.54
Q4	685	21.41
Cumulative	3,119	100

Notes: I create dummy variables for *quarter-of-birth* for all respondents, with Q1 as the reference category. Source: Author's elaboration of the STEP data for Kenya.

Balance Tests: Randomisation Checks and Consequences for Causal Inference

Before a more in-depth analysis that includes further controls and extended models with more assumptions, the balance test presents useful preliminary checks for systematic differences between the treated and control groups.

Table 2-10 Balance Test of the Reform Indicators on the Education and Skill Variables

	Coefficients	S. Error Robust S. Error	t	P>|t|	[95% conf. interval]	
		years_educ_act				
p1985_	0.949	0.203	4.670	0.000	0.551	1.348
		0.242	3.930	0.000	0.476	1.423
_cons	9.897	0.192	51.500	0.000	9.520	10.274
		0.233	42.450	0.000	9.440	10.354
		apvlit_c				
p1985_	18.545	4.443	4.170	0.000	9.835	27.256
		4.501	4.120	0.000	9.721	27.370
_cons	163.280	4.204	38.840	0.000	155.037	171.523
		4.266	38.270	0.000	154.916	171.645

Notes: Where, p1985_ is the reform dummy; years_educ_act, is the actual number of years of schooling; and apvlit_c, is the indicator for reading proficiency (cognitive skill).

With the reform indicator as a dummy variable of interest, with two categories each for the treated (1) and the control (0), Table 2.10 presents the result of the Balance Test showing the means and standard errors of key variables (schooling and skill) conditional on the status of the treatment (reform), this gives insights to the degree of unbalance in assignments to treatment and control (or the lack of randomisation) that determines the reform indicator, p1985_. Table 2.10 presents useful descriptive evidence for the Education (years_educ_act) and Skill (apvlit_c) across the two groups of interest, based on the reform indicator. The marginal increase in mean schooling and skill for the treated (p1985_) over the mean schooling and skill of the reference or control (_cons) suggests the reform explains substantial (exogenous) variation in schooling and skill. Besides presenting the mean of the schooling and skill of the control and the treated, Table 2.10 also shows the result of a balance test (t-test) aimed at testing if the reported difference in mean schooling and skills for the treated and control are statistically significant, which aid an assessment of the balance or randomisation in assignment to treatment and the control of the reform dummy. The evidence suggests substantial (and statistically significant) variations in schooling and skill are attributable to the 1985 reform. Specifically, the reform explains an average rise of 0.95 years of schooling; and 18.6 points rise in average PV (measure of reading proficiency). The nil p-values across the treated and the control for both the schooling and skill measures indicate the need to reject the null hypotheses. Hence, the variation in schooling and skill attributable to the reform are statistically significant and different from nil. Put together, this evidence suggests, the assignment to treatment and control is unbalanced (not randomised) with two out of two pairwise comparisons different (comparing schooling and skill across the treated and the control), having under 1% level of statistical significance (see Morgan and Winship, 2015; and Bellemare, 2020). This is expected for observational datasets. The issue of randomisation in assignment for schooling and skill as earlier highlighted inhibits causal inference from estimates, warranting a causal identification strategy. In addition to the evidence of (exogenous) variation in schooling and skill, further analysis conducted (see Results and Discussions) suggests the impact of the reform on schooling shows a strong dependence on background characteristics. These make the reform suitable for implementing the Difference-in-Differences and Instrumental Variables[57] approach to drawing a causal inference on the effects of background characteristics on schooling and

[57] The exogenous variation in schooling and skill attributable to the reform make the reform dummy useful for the Instrumental Variables (IV) approach used to draw causal inference from the effect of schooling on skill (research question 2). Here, schooling is the treatment that impact the outcome, skill, and does so only through the effects of the reform. Hence, it is argued that the reform only raised skills through schooling. The IV approach, although implemented in this study in responding to research question 2, it is discussed more extensively in Chapter 3.

skills; and the impacts of schooling on skill, respectively. Doing these provides useful responses to the first and second research questions respectively. Please, see the identification strategy in the Empirical Framework (next subsection) for further discussions relating to the Balance Test. Finally, the differences in the reported errors using robust standard errors instead of standard errors show the prevalence of heteroscedasticity in estimates, this suggests the need for useful controls and clustering at appropriate levels required in drawing inferences in estimates from samples to populations of interest.

2.2.3 Empirical Framework: Estimation and Identification Strategy

After discussing the theoretical framework and the data used to test the testable predictions (hypothesis) of this study[58], I now turn to the empirical framework that set the scene for this analysis by providing a research design that gives useful responses to the research questions raised.

The empirical framework is subdivided into two related parts, the estimation strategy and; the identification strategy.

Educational Attainment and Skill Determination: The Estimation Strategy

In responding to the overarching research question of this study, I deploy an estimation strategy that obtains results parametrically in reduced forms. Equations 2-5 and 2-6 are generalised models estimated using Ordinary Least Squares, for schooling and skill outcomes respectively. I estimate several specifications with some extensions (see identification strategy) of Equations 2.5 and 2.6 presenting and discussing outputs in response to the research questions raised.

Equation 2-5

$$s_i = \alpha + \gamma_{p_{1985}} p_{1985} + \gamma_B \boldsymbol{B_i} + \gamma_X \boldsymbol{X_i} + \epsilon_{ij}$$

Equation 2-6

$$S_i = \pi + \beta_s s_i + \beta_{p_{1985}} p_{1985} + \beta_B \boldsymbol{B_i} + \beta_X \boldsymbol{X_i} + e_{ij}$$

Here, s_i and S_i are the outcome variables, (years_educ_act) and skill (zapvlit_c) respectively for Equations 2-5 and 2-6. It is important to take note of schooling, s_i that is also an input

---

[58] The objective of this study entails examining the extent to which schooling impacts skill and how background characteristics and the related reform play a role in the relationship between schooling and skill in a developing context, urban Kenya to be precise.

(or independent) variable in Equation 2.6 and the parameter β_s that captures the effect of schooling on skill; P_{1985} is the reform dummy (or reform indicator); and $\beta_{p_{1985}}$ is the parameter that captures the effect of the reform. $\boldsymbol{B_i}$, is a vector of background characteristics for an individual, i, in urban Kenya, and β_B captures the effect of this. $\boldsymbol{B_i}$ consists of dummy variables indicating having a father with post-secondary schooling (father_educ_456); having a mother with post-secondary schooling (mother_educ_456); an indicator of low socioeconomic status at age 15 (ses_1); and an indicator of high socioeconomic status at the age of 15 (ses_3). $\boldsymbol{X_i}$ is a vector that captures all other covariates and β_X captures the effects of the covariates. $\boldsymbol{X_i}$ includes a wide range of controls, such as the age of the respondent that also enters the model as a quadratic term, age2. In addition to the age variable, other controls include the number of months of work experience, and tenure which also enters the model as a quadratic term, tenure_squared. Next, in this analysis, the z_j[59] are covariates that capture some district-level characteristics (number of households in the district of residence of the respondent). h_j is the measure of average schooling or/and skill in the district. It is derived by innovating using the district variable, by assigning the average years of schooling or average reading proficiency (PVs) of the district of residence of the respondent (please, see the Data Section of Chapter 4 for the specifications of these variables). π, α are the intercept terms; and e_{ij}, ϵ_{ij} are the error terms.

For both models, I test the following $(\mathrm{H_0} : \gamma = 0;\ \mathrm{H_A} : \gamma \neq 0)$ and $(\mathrm{H_0} : \beta = 0;\ \mathrm{H_A} : \beta \neq 0)$ to show the (null) hypotheses, hence, showing that the relationship between each of the outcomes (schooling and skill) and the associated inputs as shown (in Equations 2.5 and 2.6) do not statistically significantly differ from nil. Particularly, in this chapter, the overarching goal is to examine β_s (the effect of schooling on skill), showing how, γ_B (background characteristics) and the related effects of $\gamma_{p_{1985}}$ (the reform) on schooling impact β_s (the effects of schooling on skill). In other words, the goal entails examining the effects of schooling on skill, assessing how background characteristics (and related reforms) play a role in the relationship between schooling and skill, in urban Kenya. With a further aim to draw causal inference from estimates of β_s and γ_B, issues of endogeneity, particularly, unobserved factors or variables correlated with schooling, s_i and background characteristics, $\boldsymbol{B_i}$ in Ordinary Least Squares (OLS) models (see further discussions of these in the next subsection, Identification Strategy) remains a challenge to drawing causal inference from

[59] Based on World Bank's STEP design, Kenya has four strata, hence, cities are categorised as thus: Nairobi; Cities with more than 100 000 Households; Cities with under 100 000 Households but above 60 000 Households; Lastly, other Cities, with under 60 000 Households.

estimates of β_s and γ_B. Hence, the Identification Strategy that involve the use of the reform indicator, $\gamma_{p_{1985}}$ (and other Instrumental Variables) that cause exogenous variation in schooling and background characteristics, aimed at obtaining estimates of β_s and γ_B (as LATE and ATET, discussed above) from which causal inference is drawn. However, prior to the implementation of the Identification Strategy, this study commences from simple descriptive/baseline analysis from which useful insights are drawn. I now turn to discuss these descriptive analyses.

Estimating the baseline specifications of equations 2-5 and 2-6 gives useful insights that improve understanding of the problems and aid in unravelling further insights based on the research questions. Although causality is not inferred in these preliminary/descriptive analyses, I take useful steps to ensure estimates are (fairly) unbiased. Equations 2-5 and 2-6 aid the following analysis. Firstly, the relevance of the reform as a possible instrument for schooling and skill is examined. This is an initial appraisal of the reform, it entails assessing exogenous variation in schooling and skill, attributable to the reform. Secondly, I examine intergenerational (correlation and elasticity) mobility, examining how background characteristics in the forms of parental education and socioeconomic status impact the schooling and skill of the respondents (in this study, $\hat{\beta}_{B\ and}\,\hat{\gamma}_B$, are the average intergenerational regression coefficients, see models 2-5 and 2-6). These intergenerational regression coefficients capture elasticity, hence, the extent to which parental background characteristics (measures of parental schooling and wealth) impact the respondents, which explains the degree of persistence ($\hat{\beta}_{B\ and}\,\hat{\gamma}_B$) or mobility ($for\ mobility, the\ related\ parameters\ are, 1-\hat{\beta}_{B\ and}\ 1-\hat{\gamma}_B$) particularly in education. Here persistence or mobility in education relates to intergenerational changes in levels of education attained between parents and their offspring (in this case, the respondents). Besides the intergenerational regression coefficients, the correlation coefficient ($\hat{\rho}$) between parental schooling and those of their wards gives insight into the extent to which observed dispersion in the schooling of the ward (respondents) is explained by the schooling of their parents. Although correlation and regression coefficients are generally (deemed) alternatives having similar implications to mobility and persistence. However, in this study, the emphasis is on the latter (intergenerational regression coefficient). Equation 2.7 shows the relationship between the regression and correlation coefficients below, as thus:

Equation 2-7

$$\hat{\beta}_B = \frac{\sigma_{pw}}{\sigma_p^2} = \rho_{pw}\frac{\sigma_w}{\sigma_p}\text{; Therefore, } \rho_{pw} = \beta_B \frac{\sigma^p}{\sigma^w}$$

The correlation between the schooling of the parent and those of their ward (ρ_{pw}) equals the product of the intergenerational regression coefficient (β_B); and the ratio of the standard deviation of parental schooling and the standard deviation of the schooling/education of the wards ($\frac{\sigma^p}{\sigma^w}$). Equation 2.7 shows the useful relationship between the correlation and the elasticity of ‘parental capital’ (this term is taken to mean or include a measure of parental wealth as SES and parental schooling) and the human capital of the respondents. The study of Checchi et al (2013) suggests the correlation between parents and their offspring over time may also capture the effects of schooling reforms that impact education. It is well documented (in this literature) that changes in inequality in education over time strongly suggest that intergenerational correlation and elasticity evolve differently, which may make examining both measures (of intergenerational elasticity and correlation) interesting as the latter may unravel useful insights. However, due to the scope of this study, I examine the intergenerational elasticity (hence, persistence or mobility) and not intergenerational correlation. However, to better examine inequality in schooling and skills, in this study, I explore the Oaxaca-Blinder decomposition technique (see subsequent paragraphs). Central to the objective of this study, is the focus on examination of the efficiency (or productivity) of schooling by using variants of Equation 2.6. This entails examining the extent to which schooling yields skill. Equation 2.8 presents the OLS model specification for this analysis.

Equation 2-8

$$S_i = \xi + \beta s_i + \beta_1(s_i * ISCED1_i) + \beta_2(s_i * ISCED2_i) + \beta_3(s_i * ISCED34A_i) + \beta_4(s_i * ISCED4B_i) + \beta_5(s_i * ISCED56_i) + \delta_1 \boldsymbol{J}_i + \delta_2 \boldsymbol{B}_i + \delta_3 \boldsymbol{I}_i + \pi_{i\mathrm{j}}$$

As earlier noted, S_i is the standardised measure of reading proficiency (skill) which is the outcome of this OLS model. While β captures the effect of the continuous measure of schooling s_i, in this case, the parameter (β) captures the base skill level that accrues to all regardless of educational attainment. The variables $ISCED1, ISCED2, ISCED34A, ISCED4B, and\ ISCED56$ are the dummies for each of the categorical measures of schooling. $(s_i * ISCED_i)$ is the product term of continuous and categorical schooling measures and each of the parameters, $\beta_1 - \beta_5$, capture the distinct effects of schooling at the respective categories (of schooling) giving useful insights into the efficiency of schooling, inspired by Lee and Wie (2017). This model controls for several variables, with $\boldsymbol{I}_i$ a vector of individual characteristics that include the age of respondents that enters as a quadratic term; and a dummy variable for gender, with female (1) and male (0). Accounted for is $\boldsymbol{B}_i$, a vector of background characteristics and includes measures of socioeconomic status, with ses_1 indicating low socioeconomic status and ses_3 indicating high socioeconomic status; other background characteristics considered are parental education, with father_educ_456 and mother_educ_456 indicating having a father and mother with post-secondary schooling. $\boldsymbol{J}_i$, is a vector of job-related characteristics, which includes work experience and its quadratic term, as tenure and tenure_squared respectively. Other covariates include district (size or average number of households in districts) in dummies. $\pi_{i\mathrm{j}}$, is an error term; ξ in the constant term. I will now discuss the Oaxaca-Blinder decomposition technique used extensively in examining inequality in schooling and skill.

The Oaxaca-Blinder Decomposition and Applications in this study (applied in Chapters 2 and 3).

Using the Oaxaca-Blinder decomposition method, I examine heterogeneity in cognitive skills (as reading proficiency or ‘skill’ subsequently) across subsamples or groups of interest in this study. The Oaxaca-Blinder decomposition helps to improve understanding of the effects of differences in background characteristics on the variability in skill. Particularly, the decomposition gives useful insights into the extent to which variability in skill is driven by endowment; and potential discrimination attributable to background characteristics such as parental education and wealth (or ‘parental capital’). From the seminal work of Oaxaca (1973); and Blinder (1973), the Oaxaca-Blinder decomposition method has been used extensively in several studies, particularly, across the Social and Medical Sciences (Kline, 2011; Cattaneo and Wolter, 2015; Laborda et al., 2019; Rahimi and Nazari, 2021; Laible and Brenzel, 2021). As in several existing related studies, in this study, I deploy the Oaxaca-Blinder decomposition method in examining inequality (including inequity deemed an extreme form of unfair inequality rooted in discrimination, as used in this study). Several other related studies in the field of Applied Economics that explored similar decomposition techniques to examine issues of inequality are the studies of Neal and Johnson (1996); Eeckhout et al., (2014); Card et al., (2018); Brot-Goldberg et al., (2017). Lee and Wie (2017) explored several decomposition techniques in explaining the skills gap between Japan and South Korea, using the PIAAC. I now discuss how the Oaxaca-Blinder decomposition method is applied in this study.

With a continuous outcome, S_i for skill, the method entails assessing the mean difference in skill between two groups of interest[60] which are ultimately decomposed into unexplained and explained components. The latter (explained component) gives insights into the effects of differences in characteristics or endowment on the variability of the outcome (skill). The former (unexplained components) gives insight into how the other factors other than endowment impact variability in the outcome. This Oaxaca-Blinder method entails a hybrid of simple descriptive analyses — such as a first-stage multiple regression and some t-tests. I now present the methods, showing relationships in functional (mathematical) forms. With the continuous measure of skill (S_i) as the dependent variable, with n explanatory variables, ranging from $x_1, \ldots. x_n$ in a multiple regression model. The expected (mean) outcome for each of the two groups (e.g., based on the father’s education). Hence, the two subgroups in this case are (1) respondents that have fathers with post-secondary schooling; and (2) respondents that have fathers without post-secondary schooling) is given as thus:

Equation 2-9

$$\overline{S^f} = \beta_0^f + \sum_{j=1}^{n} \beta_j^f \overline{x_j^f}$$

Notice that instead of i that denotes individual-level variables, I have used j to denote group-level variables in Equation 2.9. $\bar{x}$ is the mean of each of the explanatory variables in the

---

60 In this study, groups of subsamples of interest are categorised based on background characteristics, ‘parental capital’ and includes parental education and wealth.

model. β is the estimated coefficient of the regression. However, β_0 is the constant of the regression that also accounts for the unobserved variation, hence, this includes part of the unexplained variations in the outcome. Equation 2.9 shows the mean predicted outcome (for each of the two groups of interest) as a function of the parameters/variables of interest as defined below, in this analysis, the difference in the mean of the predicted outcomes between both groups (1) and (2) is of interest, hence, given as thus:

Equation 2-10

$$\Delta\overline{\boldsymbol{S}} = \left(\boldsymbol{\beta}_0^1 - \boldsymbol{\beta}_0^2\right) + \sum_{j=1}^{n}(\boldsymbol{\beta}_j^1\overline{\boldsymbol{x}_j^1} - \boldsymbol{\beta}_j^2\overline{\boldsymbol{x}_j^2})$$

Equation 2.10 gives the effect of the difference in the mean predicted outcome, $\Delta\overline{\boldsymbol{S}}$ between the two groups, expressed as the sum of the effect of the components — here, the explained and unexplained components as earlier described constitute subcomponents that include, the coefficient differential between the two groups, as the difference in β_j. Although partly a consequence of the characteristics observed, this subcomponent is deemed an unjustified difference, hence within the unexplained component. Next is the characteristic/endowment differential between the two groups, as the difference in $\overline{x_j}$, this subcomponent is within the explained component. Lastly, the effects of the difference in the unobserved factors or variables, as $\beta_0^{\mathbf{1}} - \beta_0^{\mathbf{2}}$, this subcomponent is within the unexplained component. The Oaxaca-Blinder decomposition methodology uses a counterfactual approach in assessing the magnitude of each of the (main) components—as coefficient (or unexplained) and endowment (or explained) components—that result in variation in the outcome ($\Delta\overline{\boldsymbol{S}}$) of the model. This entails making one of the two groups a reference category that forms the basis for assessing the magnitude of the coefficient and endowment effects of the other group. Hence, the model for (2)—having a father without post-secondary education is expressed with respect to (1)—having a father with post-secondary education, as thus:

Equation 2-11

$$\overline{\boldsymbol{S}^2} = \boldsymbol{\beta}_0^2 + \sum_{j=1}^{n}\boldsymbol{\beta}_j^2\overline{\boldsymbol{x}_j^2} = \boldsymbol{\beta}_0^2 + \sum_{j=1}^{n}[\boldsymbol{\beta}_j^1 + (\boldsymbol{\beta}_j^2 - \boldsymbol{\beta}_j^1)]\sum_{j=1}^{n}[\overline{\boldsymbol{x}_j^1} + (\overline{\boldsymbol{x}_j^2} - \overline{\boldsymbol{x}_j^1})]$$

$$= \beta_0^2 + \sum_{j=1}^{n}\beta_j^1\overline{x_j^1} + \sum_{j=1}^{n}\beta_j^1(\overline{x_j^2} - \overline{x_j^1}) + \sum_{j=1}^{n}\overline{x_j^1}(\beta_j^2 - \beta_j^1) + \sum_{j=1}^{n}(\overline{x_j^2} - \overline{x_j^1})(\beta_j^2 - \beta_j^1)$$

Equation 2.11 involves $\overline{x_j^2} = \overline{x_j^1} + (\overline{x_j^2} - \overline{x_j^1})$; and $\beta_j^2 = \beta_j^1 + (\beta_j^2 - \beta_j^1)$, similarly, substituting these in Equation 2.10 yields the full specification (Equation 2.12) of the Oaxaca-Blinder decomposition.

Equation 2-12

$$\Delta\bar{S} = (\beta_0^1 - \beta_0^2) + \sum_{j=1}^{n} \beta_j^1\left(\overline{x_j^2} - \overline{x_j^1}\right) + \sum_{j=1}^{n} \overline{x_j^1}(\beta_j^2 - \beta_j^1) + \sum_{j=1}^{n} (\overline{x_j^2} - \overline{x_j^1})(\beta_j^2 - \beta_j^1)$$

In Equation 2.12, the outcome relates to (2) relative to (1). Hence, this relates to the effects of having a father without post-secondary education, relative to having a father with post-secondary schooling. Particularly, the first component $(\beta_0^2 - \beta_0^1)$ adds to the coefficient (or return) effects. It includes the unobserved (unexplained) factors that may explain potential discrimination between the groups in question; the second component $\sum_{j=1}^{n} \beta_j^1\left(\overline{x_j^2} - \overline{x_j^1}\right)$ captures the effects of the level of characteristics/endowment of the group (2) (those that have fathers without post-secondary schooling) for the group (1) (those that have fathers with post-secondary education), this term is the only term that makes the explained component as earlier discussed, rather than discriminating, the effect of this component highlights mere inequality attributable to the differences in characteristics/endowment between the advantaged and disadvantaged; the third component $\sum_{j=1}^{n} \overline{x_j^{\mathbf{1}}}(\beta_j^2 - \beta_j^1)$ entails variation in (1)'s return if they have only attained the return level of (2), as earlier highlighted, this differential return (coefficient) effect is not explained by mere variation in the characteristic/endowment between the two groups. This captures considerably, the effects of discrimination between both groups in question. The last component, $\sum_{j=1}^{n}(\overline{x_j^{\mathbf{2}}} - \overline{x_j^{\mathbf{1}}})(\beta_j^2 - \beta_j^1)$ indicates the possible interaction of the characteristic and endowment effects and relates to (or infers) the joint effects of the second and third components of Equation 2.12. The first and the third terms are typically combined to define variations in the outcome deemed unexplained (or unobserved) and consequently, the differences in coefficients (or returns) are not unjustified. Hence, these, give a measure of inequity attributable to the background characteristics between two groups. The second term defines variation in outcome deemed explained (or observed), hence, this relates to a measure of inequality as effects are due to justifiable differences in characteristics/endowment attributable to differences in background characteristics between the groups.

The Oaxaca-Blinder decomposition is not without defects. Although, it provides a useful examination of the inequality/inequity between two groups. However, at best, the decomposition provides estimates of mean outcomes for the groups of interest (depending on the reference group or category). This does not tell the extent to which distributions of the variables driving the mean outcomes are different between the groups of interest, this would be useful in understanding the mechanisms driving the mean differences. Secondly, the choice of the reference group may mar outcomes as groups of interest may not be easily comparable especially if there are several other factors (impacting a group) that are not accounted for, which may result in both selectivity and measurement issues due to possible systematic differences between the groups of interest – however, this is not the case of the groups of interest in this analysis. These biases, and the effects of omitted variables bias, can potentially result in substantial biases in the mean outcomes of decomposition, particularly, due to the unexplained and explained components of the model e.g., the effects of not accounting for a crucial variable may overstate the intercept term and understate effects of the explained component. Amidst these inherent defects of the method that are deemed inherent in Least Squares estimations and indeed in applied research of this sort, the Oaxaca-Blinder decomposition method helps to improve understanding of the inequality (and inequity) that exist between carefully selected groups like those of this study marked with clear differences in measures of background characteristics (parental education and wealth). This method is applied to improve understanding. In this study, specifically, the Oaxaca-Blinder method decomposes inequality between the groups of interest, by differences in endowment from parents to offspring; and differences in measure not accounted for by differences in endowment between the groups of interest.

The differences in characteristics (or endowment) suggest evidence of inequality. The effects of variations in the outcome unaccounted for by specific background characteristics or endowments are deemed to explain ‘potential discrimination’ between the groups of interest. Hence, such an unjustifiable difference in coefficient goes beyond mere ‘inequality’ but suggests possible effects of ‘inequity’ between the two groups of interest. Such inequity is of crucial policy relevance. Finally, the Oaxaca-Blinder decomposition also accounts for variations in the outcome (skill) attributable to factors that constitute interaction between endowment and coefficient (or returns) for the (two) groups of interest. The effects of this interaction are (mainly) interpreted conceptually. The interpretation of variations in skill attributable to the interaction terms is dependent on several factors, these include the nature of the groups in question and arguments (or concepts) typically discussed (in the literature) with the understanding of the nature of the groups.

Statistical Inference

So far, these (baseline) analyses of the effects of the reform and background characteristics on education and skill; and analyses of the impacts of education on skill are conducted without assuming causality. However, so far, I have taken steps to improve the reliability of estimates (of equations 2-5 and 2-6). A few steps include using cluster-robust standard error, ascertaining estimates are robust to heteroscedasticity. All clustering for the baseline/descriptive analyses is done at the district level, taking account of the three-stage stratified sampling technique of the STEP dataset. The sampling of the STEP data was done across four strata based on the sizes of the districts. Clustering at the district level makes it possible to draw inferences or conclusions that apply to the broader population in Kenya (see Abadie et al. 2017). Kenya is deemed to be more rural than urban and drawing inferences (from estimates) applicable to the whole of Kenya, is useful in this case. However, I interpret findings with care and some conservatism, highlighting estimates to relate to urban Kenya in most cases. However, extending Equations 2-5 and 2-6 to achieve the requirement of the identification strategy (see the following subsection/paragraph on the identification strategy), here the basis of clustering is on the treatment and not the sampling, following the work of Abadie et al. (2017). Since the treatment in this study is at the individual level, I operationalise the 2SLS-IV approach without clustering. However, the simple Difference-in-Differences approach that relies on the OLS still requires clustering at the district level to account for the sampling method used in collecting the STEP data. Lastly, the STEP data used in this study was carefully weighted to bring the main analytical sample and related subsamples to those of their respective populations. Hence, no further sample weighting was used to reach conclusions that apply to the population of interest from the samples (see Solon et al. 2015).

I will now discuss the Identification Strategy, from which causal inference is drawn from estimates using Equations 2-5 and 2-6.

Educational Attainment and Skill Determination: The Causal Identification Strategy

Finally, having evidence of an exogenous variation in education and skill; and a balance test (see Data Section) that suggests the assignment of observations to control and treatment groups are (fairly)[61] random (with some issues of heteroscedasticity which impacts the statistical significance of estimates). To draw causal inference from estimates involving Equations 2-5 and 2-6, I implement the Difference-in-Differences approach to support causal inference from the effects of parental background characteristics (as the basis of treatments

[61] Inherent to observational data, randomness in assignment to treatment and control is seldom (as opposed to experimental data). With randomness in assignments to treatment and control, estimates are causally identified, and there will be minimal requirement for an identification strategy.

on the observations) on education and skills. The Difference-in-Differences approach accounts for possible unobserved characteristics/factors that may influence schooling and skills outcomes for respondents with different parental background characteristics (basis of treatment). These unobserved characteristics are not accounted for by merely implementing the OLS specifications of Equations 2-5 and 2-6 without including the interaction terms of the reforms and background characteristics which prevents making causal statements on estimates from the equations. This approach is (particularly) useful in this situation where a lack of panel data means, there are no fixed-effect estimators that may eliminate the effects of unobserved confounders, if any. The simple Difference-in-Differences approach mitigates the bias caused by not accounting for possible unobserved characteristics using the interaction terms (see Angrist and Pischke, 2009). The Instrumental Variables (IV) approach has been deployed in this Chapter in response to research question 2. However, the IV method is discussed in detail in the subsequent Chapter where it is used more extensively.

I will now discuss the Difference-in-Differences (Diff-in-Diff) approach implemented in this study.

Simple Difference-in-Differences

If $D = 1$ indicates those that have fathers with post-secondary education (referred to as the treated units); then, $D = 0$ indicates those that have fathers without post-secondary education (referred to as the control units). Let $T = 1$ indicate the post-reform (or post-treatment) period; then $T = 0$ will indicate the pre-reform (or pre-treatment) period.

If $Y_{1i}(t)$ is the mean outcome for a respondent, i, in period t if treated before t. This is the potential mean outcome of a respondent having a father with post-secondary education before the reform implementation. Then $Y_{0\,i}(t)$ is the potential mean outcome for a respondent, i in period t if not treated before t. This is the mean outcome of a respondent who has a father without post-secondary education, before implementation of the reform. Therefore, the treatment effect (for a respondent i, at a time t,) is given as thus:

$Y_{1\,i}(t) - Y_{0\,i}(t)$.

Hence, the observed outcome is given as:

$$Y_i(t) = Y_{0\,i}(t)\,(1 - D_i(t)) + Y_{1\,i}(t)D_i(t).$$

Since the treatment (reform) takes place beyond t=0, hence, $D_i = D_i(1)$. Therefore, at pre-reform, $Y_i(0) = Y_{0\,i}(0)$; at post-reform, $Y_i(1) = Y_{0\,i}(1)\,(1 - D_i) + Y_{1\,i}(1)D_i$.

Tentatively, the average treatment effect (effect of the reform) on the treated (those with fathers that have post-secondary education) as thus:

$$E[Y_1\,(1) - Y_0\,(1)|D = 1].$$

Conditioned on a common or parallel trend of the treated and the non-treated in the pre-reform era (or with the absence of the reform), therefore,

$$E[Y_0\,(1) - Y_0\,(0)|D = 1] = E[Y_0\,(1) - Y_0\,(1)|D = 0].$$

Then, the Average Treatment Effect on the Treated is thus:

$$\big[E[Y(1)|D = 1] - E[Y(1)|D = 0]\big] - \big[E[Y(0)|D = 1] - E[Y(0)|D = 0]\big]$$

Having discussed the setup for this ‘simple’ Difference-in-Differences (DiD) approach, I now turn to specify the models estimated. Using Ordinary Least Squares (OLS), I implement Equations 2-13 and 2-14 (as above), which are simple extensions of equations 2-5 and 2-6 (which are the basis of the baseline models and estimations). Doing this helps to draw causal inferences on estimates of the (on education and skill) reform-affected background characteristics. I now turn to present the models implemented.

Equation 2-13

$$s_i = \epsilon + \beta_{p_{1985}} p_{1985} + \beta_{p1985B}(\,p_{1985} * \boldsymbol{B_i}) + \beta_B \boldsymbol{B_i} + \alpha_i$$

Equation 2-14

$$S_i = e + \gamma_{p_{1985}} p_{1985} + \gamma_{1985B}(p_{1985} * \boldsymbol{B_i}) + \gamma_B \boldsymbol{B_i} + \sigma_i$$

The terms (variables and parameters) of equations 2-13 and 2-14, are as discussed for equations 2-5 and 2-6, as below. With $(p_{1985} * \boldsymbol{B_i})$ and $\beta_{1985B,}\, \gamma_{1985B}$ (the DID estimand) in the former (equations 2-13 and 2-14) which captures the joint (interaction) effects of the reform and each of the background characteristics and gives the Average Treatment Effects (ATE). Hence this gives an estimate of the effect of the treatment on schooling and skill, on which causal inferences are drawn. The simple DiD specification is operationalised by Ordinary Least Squares (OLS) and causal identification from the DiD estimand (in

Equations 2-13 and 2-14) is subject to the regression assumptions[62]. The inherent threat of a non-parallel trend in the DiD approach results in not meeting at least one of the regression assumptions. Hence, for causal identification from the DiD approach, the effort is devoted to justifying the Parallel Trend in this study. I now briefly discuss the parallel trend and the no anticipation assumptions that must be in place for the DiD estimand to give the ATET as earlier discussed. Firstly, I explain the parallel trend in this instance. Also discussed are the simple tests, for parallel trends.

The parallel trend assumption entails, assuming (or taking steps to ensure that) before the treatment (or the reform), the **treated**, and the **non-treated** move in parallel (or in a given direction). In this analysis, this would mean, showing the effects (on schooling and/or skill) of having high socioeconomic status (or parents with post-secondary education) as the treated; and having low socioeconomic status (or parents with no post-secondary schooling) as control, are in parallel (or show a given trend) pre-reform (pretreatment).

To test or show evidence of parallel trends. In the literature, it is typical to attempt to justify the existence of parallel trends conceptually, justifying the validity of estimates from the Difference-in-Differences approach. Showing the existence of the parallel trend conceptually entails the following:

With the impact of the reform as the basis of treatment at specific times; and using parental education or socioeconomic status as the basis of the treatment in defining the treated and the control. Particularly, having high parental education and socioeconomic status classifies respondents (as treated); and those with low parental education and socioeconomic status (as control). Parents are known to maintain ‘status’ for their wards over time. Hence, it is well known that parents tend to commit to having their wards attain at least the same educational level as they did. This idea is well-founded in the literature as a mechanism of intergenerational educational mobility. The evidence from this analysis suggests a parallel trend across the treated and the control in this study. The parallel trend may be justified conceptually as one should expect a clear pattern in the schooling and skill of the advantaged relative to the disadvantaged (as the evidence from intergenerational transmission mechanisms also suggests this). However, again, the single cross-section of data means trend

[62] See Wooldridge (2015) for the detailed regression (OLS) assumptions. These, generally entail the following assumptions: All relationships are linear; All observations are independent; Nil perfect collinearity exists and non-zero variances of independent variables; The expected value of the error term is nil with any given values of independent variables; Given any values of independent variables, the variances of the error terms are equal; the error terms are normally distributed.

(over time) across the treated and control cannot be easily shown non-parametrically. In summary, I use post-estimation tests to show (visually) trends pre-and post-treatment. See post-estimation diagnostics in the Appendices (A4).

Summary of the Empirical Framework—Estimation and Identification Strategy

For ease of exposition, the next subsection responds to the overarching research questions – the effects of schooling on skills, assessing background characteristics as a mechanism through which schooling impacts skill. I will now outline the steps of the Results and Discussions subsection. I present and discuss the outputs of the analytical models in the Empirical Framework following the order of the research questions as outlined in the Introduction. Firstly, this includes examining the effects of background characteristics on schooling and skill. In responding to the first question, I analyse the effects of background characteristics on schooling and skill in an intergenerational framework (see the theoretical framework subsection). Beyond evidence from the intergenerational regression coefficient that provides useful measures of persistence or mobility in education between parents and their wards, this improves understanding of social mobility in urban Kenya. To Further obtain estimates of the intergenerational regression coefficients from which causal inferences are drawn, I implement the Difference-in-Differences technique as discussed in the Identification Strategy subsection. For the second (core) research question on the effects of schooling on skills, I examine the impacts of schooling on skills, across credential categories. Firstly, this aids assessment of the efficiency of schooling in urban Kenya. To assess heterogeneity across background characteristics, I implement the Oaxaca-Blinder decomposition technique described in the Estimation Strategy. This gives insights into the place of background characteristics as determinants or mechanisms through which schooling impacts skills. Furthermore, to obtain causal estimates of the effects of schooling on skill, I implement the Instrumental Variables (IV) technique (discussed in the subsequent chapter), exploring treatment heterogeneity by splitting the main analytical sample based on background characteristics. This aids further examination of background characteristics as mechanisms through which schooling impacts skill in urban Kenya.

Besides core parametric results that provide useful responses to the main testable predictions (research questions) as stated below, where applicable, I assess relationships of interest non-parametrically, as preliminary assessments. Together with detailed robustness checks on the main outputs/results. Attached to most of the Appendices are the results of the robustness checks. Specifically, I present the effects of the reform on schooling and skill and not merely an appraisal of the reform. This provides evidence of the exclusion restriction condition, clearly showing the impact of the reform on skill comes through the effects of the reform on schooling. Hence, showing schooling is the mechanism through which the reform impacts skill. Other robustness checks include several measures of background characteristics—such as parental education (the mother's education that is consistent with the father's education); and socioeconomic status (this includes both low and high socioeconomic status).

The objective to obtain the reform-affected estimates of the effect of background characteristics on schooling and skill together with; the need to obtain estimates of the effect of schooling on skill, from which causal inference is drawn, make the effects of reform central to this analysis. Hence, to set the scene for this analysis, I start by examining the effects of indicators of the 1985 reform on schooling and skill. As with other outputs of models implemented in this study, the first step to assessing the robustness of estimates entails careful presentation of Tables of outputs starting from the most to the least parsimonious specifications (Altonji et al. (2005)).

Using robust standard errors (clustered at the district level), I account for homoscedasticity, clustering at the district level accounts for the 3-stage stratified sampling technique used in the World Bank's STEP data. This aids useful inference from the analytical sample to the referenced population. Also accounted for, are measures of aggregate schooling (and skill) across districts (as human capital externality). Aggregate schooling enters as a quadratic. Models in this chapter account for district-level variables or factors including measures of human capital externalities. However, this is not the focus of this chapter. Hence, the effects of these controls are discussed in detail in a subsequent chapter (Chapter 4).

In the Difference-in-Differences (DiD) specification, the coefficient of the reform indicator gives an additional effect of the reform on the outcome for the untreated post-reform. The coefficient of the background characteristic captures the effect of the difference (in the outcome) between the treated and the untreated, pre-reform. Lastly, the interaction term of the reform indicator and the background characteristics gives the DiD estimand which is deemed to capture the Average Treatment Effect on the Treated (ATET), this helps to draw causal inferences on the effects (of the reform attributed to background characteristics) on education and skills.

2.3 Results and Discussions

I now present and discuss results from estimating the models presented in the empirical framework. For each research question, I begin from the outputs of the baseline (OLS) models presented in the empirical strategy for analysing outputs of the quasi-experimental approach from which causal inferences are drawn (as detailed in the identification strategy). For a start, using variants of Equations 2.5 and 2.6, Table 2.11 presents outputs of baseline (OLS) estimates of the effects of background characteristics on schooling and skill (research Q1); and the effect of schooling on skill (research Q2). Across all outputs of Table 2.11, I assess the effects of the reform indicator (p1985_) on the education and skills, of individuals.

2.3.1 Variation in Schooling and Skill Attributable to the Reform

Specifically, in Table 2.11, columns (1)-(8) present evidence of the effects of the reform (p1985_) on education, accounting for the measures of background[63] characteristics amidst other variables; and columns (9)–(16) present evidence of the effects of the reform on skill, accounting for measures of schooling amidst other variables. Where applicable, the following analysis or discussions will set the scene for further analyses in this, and the subsequent chapters.

Columns (1)–(5) show the baseline estimate of the effect of the reform on schooling is a rise of at least 0.516 years of education which is robust to the inclusion of all measures of background characteristics of interest, strata- and district-specific effects (5). Interestingly, comparing (5) and (6), the evidence suggests that the loss of the statistical significance of the reform indicator is attributable to not accounting for the indicator of the mother's post-secondary schooling and high socioeconomic status. Hence, not accounting for these, the increase in the mean of the indicators of father's post-secondary schooling; and low socioeconomic status amidst the loss of the statistical significance of the reform indicator further suggests the following. Firstly, the effects of all background characteristics (parental education and socioeconomic status) are subsumed[64] in the effect of the father's post-secondary schooling, and the indicator of low socioeconomic status. Secondly, the loss in the statistical significance of the reform indicator further suggests that the latter specification (6) provides more robust estimates of the effects of background characteristics and other

[63] Background characteristics used in this study include indicators of the father's tertiary education; mother's tertiary education; low socioeconomic status; and high socioeconomic status.

[64] This is also the case of columns (13) and (14).

inputs to the model inclusive of the reform indicator that shows a good fit. Subsequently, the focus is on the indicators of low socioeconomic status and the father’s post-secondary schooling as useful measures of background characteristics in urban Kenya. However, I explore the use of indicators of mother's post-secondary schooling; and high socioeconomic status as robustness checks on the effects of the indicators of father’s post-secondary schooling; and low socioeconomic status.

Similarly, Columns (9)-(15) show evidence of the effect of the reform on skill is at least, a standard rise of 12.1 points, statistically significant at the 5% level and robust to the inclusion of all covariates of interest in this analysis, see (15). However (16) is an exception as the effect of reform on skill, is statistically insignificant (not different from nil). The difference in the effect of the reform across both specifications ((15) and (16)) of the model may be attributable to several factors. This may include—the effects of district-level schooling; and not accounting for strata-specific effects. Particularly, the higher mean values and statistical significance of individual- and district-level schooling that is accompanied by the loss of the statistical significance of the effects of the reform in (16), suggests that the effect of the reform on skill is strongly dependent on the effect of schooling on skill—accentuating the strong moderating effects of schooling on the impact of the reform on skill. This understanding is crucial to the identification strategy followed in this study.

Furthermore, a quick review of (14)-(16) suggests that the statistical significance in the estimates of the effect of the stratum characteristics (based on district size) is adversely impacted, accounting for the measures of schooling (individual- and district-level). Particularly, this is evident in comparing (14) and (15), where accounting for schooling (district- and individual-level) in (15) increases the mean (adverse) effects of district size on skill. Interestingly, this is consistent with the effects of district size on schooling, comparing (6) and (7). This suggests that the effect of human capital (individual- and district-level schooling and skill) is highly correlated to district size as one would expect – those with high skills and schooling are attracted to large cities. Hence, (subsequently) to mitigate this effect, I use model specifications that do not simultaneously account for measures of human capital and district size. Therefore, for schooling as an outcome, instead of (7), I estimate (6) and (8) as variants of Equation 2.5; for skill outcome, instead of (15), I estimate (14) and (16) as variants of 2.6, as explained in the Empirical Framework. Model specifications (8) and (16) that account for human capital (schooling and skill respectively), indicating a relatively high coefficient of determination (r2), suggesting the models sufficiently explain variations in the outcomes, relative to (6) and (14) that account for district characteristics.

One last point to note (although discussed in more detail in a subsequent chapter) is that comparing (8) and (16) suggests that, although district-level schooling enters as a quadratic term in (16), this may not be the case for district-level skill in (8) where effects of district-level skill specified to include the quadratic term is statistically insignificant. However, this enters as a quadratic term in (7) that simultaneously accounts for district size and human capital (individual and district level skill), although, deemed spurious. However, further evidence shows that the district-level skill only enters linearly, statistically significant at the 5% level in a variant of Equation (8) (see Appendix 1 (A1) Table 1).

Besides setting the scene for this study, preliminary discussions on the effects of the reform are germane in explaining the effect of the reform on schooling and skill—which form the basis of drawing causal inferences from estimates. I will now examine the variation in schooling and skill attributable to the reform, in urban Kenya.

Table 2-11 Baseline (OLS) Estimates: Effects of the Reform and Background on Schooling and Skill

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	Education (Actual Years of Schooling), years_educ_act.								Skill (Standardised Reading Proficiency), zapvlit_c.							
reform, p1985_	1.174***	0.655*	1.104***	0.644*	0.516*	0.498	0.0940	0.406	0.237***	0.129*	0.225***	0.131*	0.186**	0.183**	0.121*	0.090
	(0.000)	(0.011)	(0.000)	(0.011)	(0.043)	(0.053)	(0.679)	(0.096)	(0.000)	(0.026)	(0.000)	(0.025)	(0.001)	(0.002)	(0.030)	(0.111)
father_educ_456		2.477***		2.166***	2.109***	2.776***	1.487***	1.716***		0.589***		0.525***	0.546***	0.612***	0.257***	0.245***
		(0.000)		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)		(0.000)		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
mother_educ_456		1.313***		1.221***	1.197***					0.117*		0.0998	0.108*			
		(0.000)		(0.000)	(0.000)					(0.023)		(0.051)	(0.032)			
ses_1 (low)			-1.804***	-1.453***	-1.442***	-1.522***	-1.074***	-1.059***			-0.331***	-0.258***	-0.257***	-0.269***	-0.0693	-0.0713
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)			(0.000)	(0.000)	(0.000)	(0.000)	(0.156)	(0.142)
ses_3 (high)			1.286***	0.463*	0.521*						0.270***	0.111*	0.0809			
			(0.000)	(0.034)	(0.016)						(0.000)	(0.023)	(0.078)			
strata_N					0.752	0.737	1.724***						-0.267**	-0.270**	-0.413***	
					(0.062)	(0.072)	(0.000)						(0.004)	(0.004)	(0.000)	
strata_L					0.653	0.642	1.383***						-0.243**	-0.245**	-0.374***	
					(0.066)	(0.074)	(0.000)						(0.004)	(0.004)	(0.000)	
strata_M					0.778*	0.755*	1.499***						-0.248***	-0.252***	-0.393***	
					(0.012)	(0.017)	(0.000)						(0.000)	(0.000)	(0.000)	
avg_skill (district)							-0.0981*	-0.00375								
							(0.013)	(0.902)								
avg_skill_sq (district)							0.00031**	0.00004								
							(0.005)	(0.648)								
skill (individual, not Z)							0.0210***	0.0202***								
							(0.000)	(0.000)								
avg_yos (district)															0.121	-0.370***
															(0.324)	(0.000)
avg_yos_sq (district)															-0.006	0.016***
															(0.325)	(0.000)
years_educ_act (indiv)															0.127***	0.128***
															(0.000)	(0.000)
_cons	9.954***	9.744***	10.29***	10.11***	9.990***	10.12***	13.60***	6.251*	-0.148*	-0.189**	-0.092	-0.128	0.130	0.149*	-0.643***	-1.476***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.023)	(0.042)	(0.007)	(0.196)	(0.064)	(0.082)	(0.043)	(0.000)	(0.000)
N	3145	3145	3145	3145	3145	3145	3145	3145	3145	3145	3145	3145	3145	3145	3145	3145
R-sq	0.009	0.130	0.071	0.159	0.166	0.157	0.364	0.342	0.006	0.080	0.040	0.094	0.109	0.107	0.102	0.291
adj. R-sq	0.009	0.129	0.070	0.157	0.164	0.155	0.362	0.340	0.005	0.079	0.039	0.093	0.106	0.105	0.100	0.290

Note: Table reports outputs of variants of models 2.5; and 2.6, as columns 1-8; and 9-16 respectively. The outcome of the former is schooling as actual years of schooling and the latter skills, as standardised reading proficiency. Both baseline models estimate the effects of the indicator of the 1985 curriculum structural reform, p1985_; and background characteristics (as indicators of the father's post-secondary education, feduc; indicator mother's post-secondary education, meduc; low socioeconomic status, ses_1; high socioeconomic status, ses_3 with ses_2 or average socioeconomic status as the reference category) on schooling and skill; and the baseline effects of schooling and skill. Additional control variables include strata-specific (district-size) effects based of the stratified sampling (based on the number of households (HH) across cities). This categorises all districts into either of the following: Nairobi, strata_N; Other Large districts, strata_L with over 100 000HH; Medium districts, strata_M with over 60 000HH but under 100 000HH; and other districts, strata_S with under 60 000HH (reference category). Other controls include average schooling, avg_yos across districts which also enters as a quadratic term, and average skill (not standardised). Other covariate includes an indicator of female gender. p-values in parentheses: * $p<0.05$; ** $p<0.01$; *** $p<0.001$. With robust standard error, clustered at the district level.

Table 2.11 shows a rise of up to 1.174 years of schooling and a standard rise of up to 23.7 percentage points in cognitive skill is attributable to the reform ((1) and (9)). These positive and statistically significant baseline estimates are at best, average (estimates) of the effects of the reform on schooling and skill, regardless of background characteristics (although accounted for background characteristics but not specific to a group of respondents based on their background characteristics). Following the specifications of the models presented in columns (6) and (14) of Table 2.11, Table 2.12 presents similar outputs (specifications) following the Difference-in-Differences (DiD) approach. The estimates from the DiD approach give useful insights into the variability of the impact of the reform based on background characteristics of interest in this study. The evidence from Table 2.12 suggests, relative to the advantaged, the reform favourably impacts the schooling (and skills) of the disadvantaged[65]. Particularly, (1) and (2) suggest that a rise of 0.695 years of schooling and a standard rise of 18.9 percentage points in cognitive skills are attributable to the effects of the reform, for the untreated (those who have fathers without post-secondary education) relative to the treated (respondents that have a father with post-secondary schooling).

Both estimates are statistically significant at the 1% level. This evidence is consistent using the indicators of the mother's post-secondary schooling ((3) and (4)); and high socioeconomic status ((7) and (8)) as the basis of treatment. The latter ((7) and (8)) show the strongest effects of the reform on the schooling and skills of the untreated (those that have mid-low socioeconomic status). This suggests relative to those with high socioeconomic status, for those with mid-low socioeconomic status, a rise in 0.96 years of schooling; and a standard rise of 28.3 percentage points in cognitive skill are attributable to the reform. These effects are statistically significant at the 0.1% level. Interestingly, using the indicator of low socioeconomic status as the basis of treatment (as in (5) and (6)) provides evidence against the argument that the effect of the reform on skill is through its effect on schooling. Using the indicator of low socioeconomic status, the evidence suggests that, relative to the treated, the untreated (those with med-high socioeconomic statuses) have no statistically significant rise in their schooling attributable to the reform (5). This is consistent with other measures of background characteristics considered (as the basis of treatment). However, evidence from (6) suggests that a standard rise of 22.5 percentage points in cognitive skills (statistically significant at the 1% level) is attributable to the reform, for the untreated. The evidence from (5) and (6) suggests no statistically significant effect on schooling but a

---

[65] The terms 'advantaged' is used to describe those that have fathers and/or mothers with post-secondary education; and/or those with high socioeconomic status at the age of 15. The term 'disadvantaged' is used to describe those that have fathers and/or mothers without post-secondary schooling.

statistically significant effect on the skill is attributable to the reform for respondents with mid- and high- socioeconomic status. This suggests that the argument (from the previous subsection) that the effects of the reform on the skill of the respondents is through the effect of the reform on their schooling may not hold for all. However, the evidence (6) of a statistically significant effect on the skill of those with high- and mid-socioeconomic status (22.5**) is not as high and strong as the effect of the reform on the skill for those with low- and mid-socioeconomic status (24.3***). Hence, the evidence from all indicators of background characteristics (as a basis of treatment) is consistent with the argument that suggests, the reform raises the schooling and the skill of the disadvantaged, relative to the advantaged.

Put together, the evidence suggests that the reform results in (exogenous) variation in schooling and skill, particularly for the disadvantaged. Suggesting that background characteristics are mechanisms through which the reform impacts schooling and skill (or simply, the effects of the reform on the education and skill of respondents are strongly dependent on their background characteristics).

To further examine the effects of the reform across categories of education (*isced*) and skill (apvlit_c). This aids further examination of the impact of the reform on schooling and skill, particularly, beyond showing the mean effects on schooling and skill, regardless of the categories of education and skill. This analysis gives insights into understanding the categories of schooling and skill most impacted by the 1985 curriculum reform—revealing the 'dynamic' impacts of the reform. A useful rationale for understanding this dynamic effect of the reform includes the following. Pre-reform, relative to the advantaged, for the disadvantaged, in addition to having minimal access to schooling, the disadvantaged tend to have certain categories of schooling that provide certain skill levels. In addition to the effects of background characteristics, the reform can make it such that those with certain background characteristics attain certain categories of schooling and therefore certain skill levels. To further examine the rise in schooling and skill attributable to the reform as discussed in the previous subsection, I deploy simple Probit models, regressing dummies of categories of schooling and skills on the reform indicator, accounting for background characteristics, strata-specific effects, following similar specification of Equations 2.5 and 2.6 as in (6) and (14) of Table 2.11. I now examine the categories of schooling and skills most impacted by the reform.

Table A1.2 (see Appendix A1) shows that, regardless of background characteristics, while the reform substantially raises the probability of attaining the isced34A[66] credential, a fall in the probability of attaining the isced56[67] credential category is attributable to the reform. Both effects are statistically significant at 0.1% and 5% levels respectively. To further examine this, I disaggregate the main analytical sample, by background characteristics. The evidence suggests, relative to those who have fathers without post-secondary education credentials, a high probability of attaining isced34A credential is attributable to the reform for those that have fathers with post-secondary education credentials ((3) and (5)). However, this is reversed as relative to those that have fathers without post-secondary education credentials, a fall in the probability of attaining isced56 credential is attributable to the reform for those that have fathers with post-secondary education credentials ((4) and (6)). Using a measure of socioeconomic status (as defined in this study), the evidence suggests, that the reform has no statistically significant effects on the probability of schooling for those with low socioeconomic status ((7) and (8)), however, consistent with the findings using parental education, the evidence suggests, the reform raise the chances of attaining the isced34A credential for those with mid-high socioeconomic status. Overall, consistent with the findings from the aggregated or main analytical sample ((1) and (2)) where evidence suggests, a rise in the probability of attaining the isced34A credential; and a fall in the probability of schooling at the isced56 credential category.

Disaggregating the main analytical sample based on background characteristics, the evidence suggests that the reform raised the chances of attaining secondary and some post-secondary schooling for all, particularly, for the ‘advantaged’. However, a fall in the probability of attaining university education for the ‘advantaged’ is also attributable to the reform. Table 3 (see Appendix A1) shows evidence of the effects of the reform on skill. Similarly, for skills, regardless of background characteristics, the evidence suggests that a rise in the probability of attaining level 3 skills is attributable to the reform. However, no statistically significant effect on level 4 skill is attributable to the reform. Disaggregating across background characteristics, the evidence suggests that, at best, the reform raises the probability of attaining level 3 skills for respondents having parents with no post-secondary education credentials, those with low socioeconomic status, and those with mid-high socioeconomic status. However, the evidence suggests no statistically significant effect on

[66] The ISCED34A credential category is for the completion of secondary and some post-secondary education (please, refer to the Data Subsection).

[67] The ISCED56 credential category is for completion of tertiary (university) education. Please, refer to the Data Section for detail.

the skills of respondents (having fathers with post-secondary schooling) is attributable to the reform. Findings on the impact of the reform on the probability of certain skill and credential categories accentuate the argument on schooling as a mechanism through which the reform impacts skill. The evidence suggests that, compared to the high credential category, the reform raises the probability of attaining the low credential category, which, in turn, raises the probability of (attaining) the low skill level. Furthermore, although using continuous schooling and skill (as outcomes), the evidence suggests, a substantial rise in the schooling and skill for the 'disadvantaged'. Further evidence from using categorical schooling and skill (as outcomes) suggests that the rise in the schooling and the skill of the disadvantaged are mainly at the low levels (non-tertiary education, and at most, level 3 skill).

So far, the findings from examining variations in continuous and categorical schooling and skill attributable to the reform provide a basis for further analysis of the main testable predictions (research questions and arguments raised) in this study[68]. I now turn to further analyse and discuss variations in schooling and skill attributable to background characteristics with emphasis on the indicators of the father's post-secondary education and low socioeconomic status as key measures of background characteristics[69].

2.3.2 Effect Background Characteristics on Schooling and Skill

Table 2.12 reports estimate of the Average Treatment Effects on the Treated (ATET) by deploying the Difference-in-Differences (DiD) technique. With the focus on the indicators of the father's post-secondary education; and low socioeconomic status, as measures of

[68] From the below findings on the effects of the reform (policy dummy) on schooling and skill, it is evident that the reform results in an (exogenous) variation in schooling and skill. See Appendix A3 for tests of instrument validity. This is also discussed in detail in the final chapter (Chapter 6). The evidence affirms that the effects of the reform on skill is (substantially) through the effects of the reform on schooling, suggesting schooling mediates the effect of the reform on skill. As earlier highlighted, this understanding is crucial to the identification strategy pursued in this study. As earlier highlighted, without this, the reform will remain an invalid instrument in the (IV) approach. Inspired by the works of Angrist and Krueger, (1991); and Bound et al. (1995), in addition to exploiting the reform as an instrument, I exploit (exogenous) variations in schooling attributable to the Quarter of Birth variable (as earlier described, see data subsection) and possible variations attributable to an interaction of both the reform dummy and the quarter of birth variable. A second point made more succinct in the analysis that show (exogenous) variations in schooling and skill attributable to the reform is the effects of background characteristics. Disaggregating across background characteristics suggests a substantial change in the effects of the reform on the schooling and skill of respondents. These strongly suggests a dependence of the effect of the reform on background characteristics. This makes it possible to provide a useful response to the set objective of this study (research question 1) by exploiting the reform as a treatment in drawing causal inferences from estimates of the 'effects of background characteristics on schooling and skill in a Difference-in-Differences approach.

[69] Besides the indicator of the father's post-secondary education and the indicator of low socioeconomic status, other measures of background characteristics considered in this study include socioeconomic status (high) and indicator of the mother's post-secondary education.

background characteristics. The following analysis is in response to the first[70] research question raised. This entails understanding the extent background characteristics promote or inhibit schooling and skill. There has been a drive for a rise in quality inputs in schools in developing countries in recent times. This analysis provides a useful basis for the argument (as raised) in support of 'equity in access to schooling' over 'quality inputs in schooling' for increased skill levels in the non-OECDs. This is particularly useful for sub-Sahara Africa characterised by problems of resource constraints in school provision. In this analysis of the effects of background characteristics on schooling and skill, the baseline (Table 2.11) output is considered the basis of the conclusions drawn in this analysis is the main output (Table 2.12) which is based on the DiD specification and gives the ATET from which causal inference is drawn. I will now turn to discuss both the baseline and DiD outputs.

At first, examining the relationships non-parametrically, the findings suggest, evidence of substantial differences in outcome (schooling) across background characteristics, particularly, substantial variations in schooling exist between those that have fathers with post-secondary schooling (or the 'advantaged') relative to those that have fathers without post-secondary schooling (or the 'disadvantaged'). Please, see further descriptive evidence in Tables A1.7.1 and A1.7.2 in Appendix A1. Fig 2-5 indicates that over fifteen percent of those who have fathers with post-secondary schooling have fifteen (15) years of schooling, however, just a little over five (5) percent of those having fathers without post-secondary schooling have fifteen years of schooling.

[70] The first research question is on the effects of background characteristics on schooling and skill in urban Kenya.

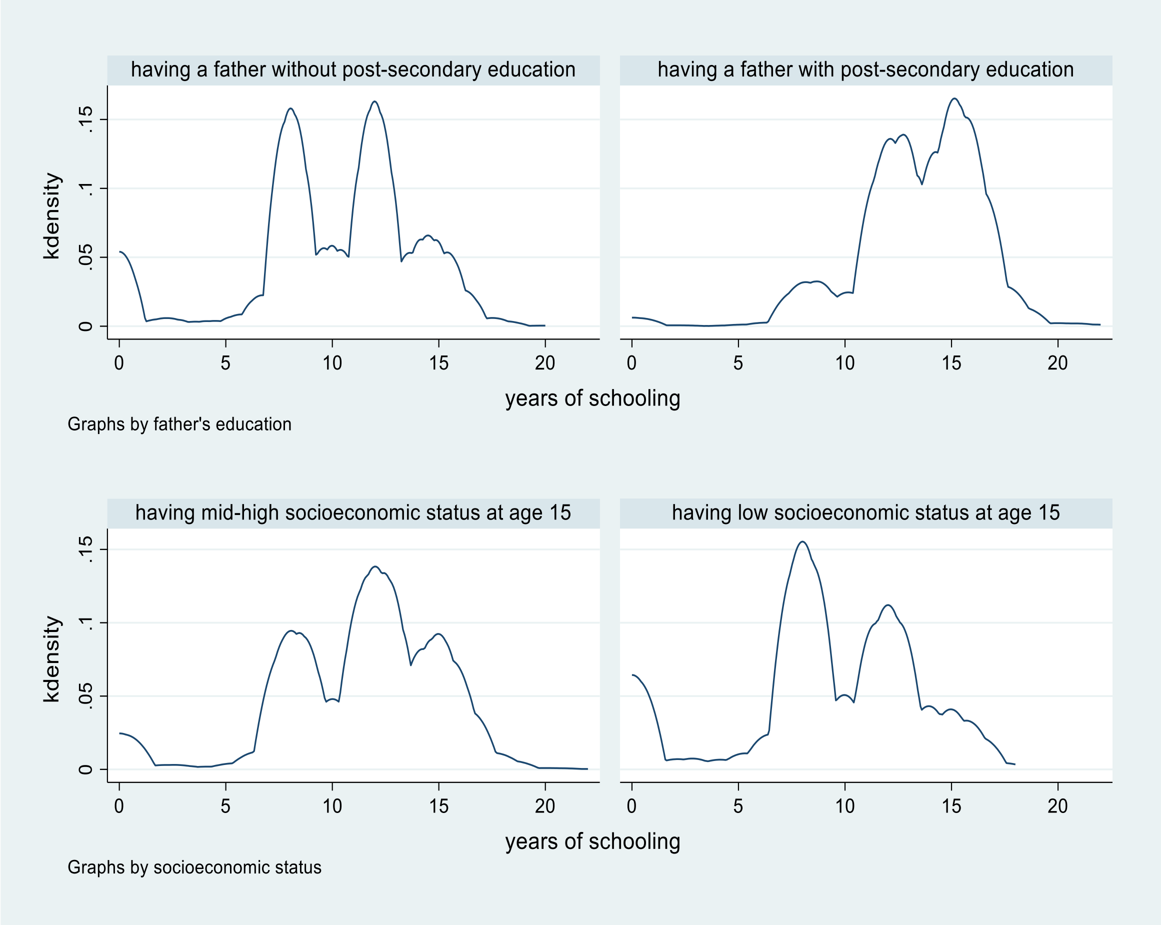


Figure 2-5: Kernel Density Plot, Years of Schooling, by Backgrounds.

Source: STEP Skills Measurement Program of the World Bank.

Besides, evidence from the graph suggests that relative to those who have fathers with post-secondary schooling, a substantial number of respondents with under two years of schooling have fathers without post-secondary schooling. A similar trend exists based on socioeconomic status (at age 15). Relative to those with mid-high socioeconomic status, those with low socioeconomic status have less schooling. These make high proportion of those with under two years of schooling.

For the rest of this analysis, I discuss the evidence or estimates from the parametric analysis—the effects of background characteristics on schooling and skill—assessing mechanisms through which defined predictors impact the outcomes. Where applicable, I discuss the robustness of estimates and highlight discussing limitations.

Effects of Background Characteristics on Education and Skill – Baseline & DiD

Baseline

Table 2.11 has outputs of Equation 2.5 that include the indicators of background characteristics (including the father's post-secondary education) as predictors of schooling. The baseline evidence suggests that the father's post-secondary schooling results in a positive and statistically significant effect on the schooling of their wards, this indicates a strong intergenerational transmission (effect) of the schooling of fathers on the schooling of their wards. Specifically, the choice specification, column (6), suggests that having a father with post-secondary education explains an additional 2.776 years of schooling for their wards. This is statistically significant at the 0.1% level and robust to (inclusion of) the strata-specific effects. Hence, the rising educational attainment across Kenya (and sub-Saharan Africa by extension) as suggested by the study of Barro and Lee (2013) is at least, partly attributable to the intergenerational effects of (parental) schooling. This finding is consistent with the effect of socioeconomic status deemed associated with parental schooling. A loss of 1.522 years of schooling of respondents is attributable to having a low socioeconomic status at age 15. Columns (9) – (16) of Table 2.11 present outputs of Equation 2.6 with skills as an outcome. Column (14) suggests a standard rise of 61.2 percentage points in reading proficiency is attributable to having a father with post-secondary schooling, again, this is consistent with the effects of socioeconomic status as a standard fall of 26.9 percentage points in reading proficiency is attributable to having a low socioeconomic status at age 15. So far, all estimates are statistically significant at the 0.1% level. However, for the baseline estimate for skill (as an outcome), the choice specification accounts for individual and district-level schooling, as in column (16) of Table 2.11. Interestingly, accounting for schooling, the evidence suggests that a father's post-secondary education is associated with a standard rise of only 24.5 percentage points in cognitive skills (or reading proficiency), statistically significant at the 0.1% level. Here, the effect of low socioeconomic status is not different from nil, hence statistically insignificant.

Comparing findings on the effects of a father's post-secondary schooling on the schooling and skill of their ward (respondent), the evidence suggests that the effects of background characteristics on skill is dependent on the effects of both individual- and district-level schooling. Relative to the estimate of the effects of father's post-secondary schooling on the skill of the respondents, as in column (14), accounting for individual and district-level schooling instead of the district- and stratum-specific effects explains a shortfall of 36.7

percentage points in the effect of father's post-secondary schooling on skill as in column (16). Furthermore, a loss of the statistical significance of low socioeconomic status is also attributable to accounting for district- and individual-level schooling. This suggests that, while background characteristics strongly impact schooling, the effect of background characteristics on skill is (strongly) dependent on the effects of schooling on skill. Hence, background characteristics (substantially) impact skill by first impacting schooling, which, in turn, impacts skill. Suggesting schooling mediates the effects of background characteristics on skill. This is also evident from Table 2.13 where the OLS effect of schooling on skill (accounting for background characteristics) is examined, in the next subsection. The evidence from this baseline analysis suggests that having a father with post-secondary education is associated with a three-year difference in schooling; and a standard rise of 24.5 percentage points in cognitive skills, accounting for schooling deemed to mediate the effects of background characteristics on skill.

Difference in Differences

Table 2.12 presents the main findings on the effects of background characteristics on schooling and skill, following the DiD technique, as earlier described (to draw causal inference). Columns (1)-(8) follow similar specifications for schooling and skill as in columns (6) and (14) of Table 2.11 for the baseline output, particularly, controlling for stratum effects with the same set of covariates for both schooling and skill outcomes. In columns (1) and (2) of Table 2.12 where the basis of the treated is the father's post-secondary schooling, the evidence suggests, pre-reform, the single difference (between the treated and the untreated) in schooling and skill is a rise of 5.4 years of schooling and standard rise of eighty-seven (87) percentage points in cognitive skill respectively, are associated with having a father with post-secondary education (pre-reform). However, post-reform, having a father with post-secondary schooling (the treated) results in a loss of an average of 2.53 years of schooling (statistically significant at the 1% level) and no statistically significant effect on skill. These estimates (the DiD estimands) capture the Average Treatment Effect (effects of the reform) on the Treated (those that have fathers with post-secondary education), this is the basis on which causal inference is drawn, hence the basis of the conclusion in this analysis. For some robustness, the DiD estimand (ATET) for schooling and skill outcomes using the indicator of the father's post-secondary education are consistent with the use of the indicator of the mother's post-secondary education (columns (3) and (4)). However, using the mother's post-secondary schooling suggests a slightly less statistically significant effect (at the 5% level) but a greater mean coefficient, relative to the use of the father's post-secondary education. Furthermore, evidence from indicators of socioeconomic status (columns (5)-(8)) shows the expected pre-reform effects on schooling and skill. However, post-reform, further evidence suggests, that no statistically significant effect on schooling and skill is attributable to socioeconomic status as measures of background characteristics. The zero or statistically insignificant effect of socioeconomic status post-reform is interesting as it indicates the ATET is nil, suggesting the reforms may have better managed (possible) inequity in schooling and skills attributable to socioeconomic status relative to those attributable to other background characteristics such as parental education. This is quite probable as reform can easily focus on the needs of certain groups based on their perceived (current) socioeconomic status[71], however, it may be less likely (relatively) to have a reform focus on groups based on their parental education. While the latter is not impossible, it is important to note socioeconomic status is highly correlated to parental education (this is a rationale for having both as measures of background characteristics in this study. This is not specific to this study as it is also typical to take parental education as a measure of socioeconomic status). Besides the need to examine the effects of each of these variables in this study, using (both) variables supports the robustness checks in this analysis.

Baseline and DiD Outputs:

Intergenerational Transmission of Schooling between Respondents and Parents

After discussing findings from the baseline and DiD outputs based on our objective (effects of background characteristics on schooling and skill), I compare the findings of the baseline and DiD outputs. This helps to better give a response to the secondary objective of this study which entails assessing intergenerational transmissions from parents to offspring. Furthermore, comparing the baseline to the DiD outputs helps to assess the robustness of key findings based on the primary objective of this study. I will now discuss the main findings of the effects of intergenerational transmissions of schooling. The evidence from the DiD output suggests, unlike the baseline (OLS) estimates, where average effects of background characteristics show robust evidence of (at least) persistence or upward mobility for the treated (those that have parents with post-secondary schooling) relative to the untreated. The evidence from this quasi-experiment suggests relative to having parents without post-secondary schooling, the effect of parental post-secondary schooling (and the

[71] The socioeconomic status variable as used in this study, indicates how wealthy the parents of respondents were when the respondents were aged 15. It classifies respondents to either, high-SES; low-SES; and mid-SES (reference). Please, see the data subsection for more.

reform) is a loss of 2.53 years of schooling. Hence, the effect of having parents without post-secondary schooling is a rise of 2.53 years of schooling. This suggests evidence of downward mobility for the treated and upward mobility for the untreated are attributable to the reform. It is important to note that these findings for the baseline and DiD outputs are distinct, although the DiD estimand gives the reform-affected estimate, the latter (baseline) estimate explains the average effect of the father's post-secondary schooling regardless of the timing of the reform, this does not mean the baseline output is completely devoid of the reform but takes into account all other possible factors (unaccounted for (omitted variable)) that may drive the effects of father's post-secondary schooling. Tentatively, both results seem plausible however, may be interpreted differently. Effectively, a father's post-secondary schooling explains an additional 2.78 years of schooling for their ward (Table 2.11 column (6)), suggesting upward mobility (or some persistence) for the treated or respondents that have fathers with post-secondary schooling. From this, we may infer that the reverse may be the case for the untreated. However, disentangling the effects of the reform (Table 2.12), we may infer that, having a father with post-secondary schooling results in a loss of 2.53 years of schooling attributable to the reform (the joint effects of the treatment for the treated, ATET). Hence, having a father without post-secondary schooling results in a 2.53 increase in years of schooling attributable to the reform. Suggesting the reform may have been aimed at managing inequality/inequity (by impacting mobility/persistence in the schooling of parents and their wards) in schooling between the advantaged and the disadvantaged as evident in subsection 2.3.1 (where the sole effect of the reform was discussed) where the disadvantaged seem to be more favourably impacted by the reform. The reform may alter the intergenerational transmission of schooling between parents and offspring.

Robustness Checks

I now compare the findings of the baseline and the DiD outputs, to further test the robustness of the effects of background characteristics on schooling and skill. Hence, although some useful consistency exists between the outputs of the baseline and the DiD estimations the mean estimates of the baseline and DiD outputs differ substantially. The latter is expected, as the mean effects of the DiD estimand (ATET) relate to the treated in the treatment category, the baseline estimate is a first approximation of the Average Treatment Effect (ATE) that gives a useful approximation of the average effects of the treated regardless of their treatment status. Almost ninety percent of the analytical sample is in the treatment category (see the specification of the 1985 reform dummy in the Data Subsection). This may drive some useful consistency in the mean estimates of the baseline and DiD outputs as earlier described. However, substantial variability in the mean estimates of the baseline; and DiD still exists. To emphasise this again, the basis of the treated category is background characteristics, and the basis of the treatment category is the reform. To further unravel some of the consistencies between the baseline and the DiD technique, I highlight (or compare) the pre- and post-reform effects of background characteristics on schooling and skill as the outputs of the quasi-experiment (DiD); and the related outputs of the baseline estimates. Hence, from the preferred (choice) specification of the baseline models for schooling and skill outcomes (see columns (6) and (16) of Table 2.11 for schooling and skill outcomes respectively), the evidence suggests that having a father with post-secondary schooling is associated with a rise in schooling of almost three years. However, the baseline output for the effect of background characteristics (parental postsecondary education) on skill is relatively small (a standard rise of 24.5 percentage points) but statistically significant at the 0.1% level, after accounting for individual and district level schooling (compare columns (14) and (16) of Table 2.11). The former (effects on schooling) is evident from the outcomes of the quasi-experiment (columns (1) and (2) of Table 2.12) as the DiD estimand suggests a loss of about 2.5 years of schooling is attributable to having a father with post-secondary schooling, post-reform and; pre-reform, estimates show a rise of about 5.5 years of schooling, netting these together show some consistency with the baseline that suggests having a father with post-secondary schooling explains a 3-year difference in schooling regardless of treatment category. However, some inconsistency between the baseline and DiD estimates exists for the effects of background characteristics (particularly, parental education) on skill as, although the baseline outcome shows positive and statistically significant effects on skill (a standard rise of 24.5 percentage points) after accounting for schooling, however, the effects from the DiD estimand (post-reform) is statistically insignificant, hence, not different from nil even with the substantial positive and statistically significant effect of parental education (standard rise of about 86.9 percentage points, statistically significant at the 0.1% level) for the DiD pre-reform estimates. Netting or averaging the pre-reform A causal inference may be drawn from the nil effect of the DiD estimand and may be compared to the 24.5 percentage points increase in the baseline outcome by arguing Omitted Variable bias in the baseline estimate – suggesting controlling for more (omitted) variables in the baseline model may effectively result in a nil effect of the parental education on skill. However, this variability in estimates of the effects of background characteristics on skills between the DiD and baseline techniques cannot be easily reconciled. Suggesting, either the baseline or DiD output may have some spurious effects. However, with a focus on the output on which causal inference may be drawn (as an objective of this analysis), outputs of Table 2.12 that deployed the DiD technique are further examined.

Further Discussions on Findings – For Further Analyses:

The current nil outcome of the effects of background characteristics (parental education) on skill in the DiD technique suggests a situation where the effect of schooling on skill is problematic (if schooling does not give skill as due, this will invalidate the inference/understanding on the effects background characteristics on skill, deemed to be mediated by schooling), in this case, the current findings using the DiD technique will be sufficiently robust as the output from the DiD accentuate the lack of skill from schooling. However, at the extreme, another possible way to explain the current nil outcome for the effects of background characteristics on skill using the DiD approach may stem from the inherent defects with the method (ATET), as it only considers the effect of the treatment on the treated which does not truly give the Average Treatment Effects (ATE) or estimates directly comparable to the baseline estimates (deemed a first approximation of ATE). For this, other useful quasi-experimental approaches may offer useful insights. Particularly, the Two-Stage Least Square Instrumental Variables (2SLS-IV) approach that gives the Local Average Treatment Effects (LATE) may provide more consistent estimates that can compare to the baseline and DiD outcomes. The LATE gives, the ATE of the treatment category (hence, the ATE specifically for those impacted by the reform) which may make it a useful robustness check on the ATET, as previously described.

Efficiency in Schooling

The Efficiency of Schooling is discussed in more detail in the subsequent subsection. However, to discuss some related findings on background characteristics, this subsection briefly discusses the Efficiency of Schooling, particularly, showing how the efficiency-in-schooling may influence background characteristics. As earlier argued, the effects of background characteristics on skill show dependence on its effects on schooling. Hence, suggesting background characteristics impact skill by first impacting schooling, and schooling then impacts skill in turn. To be more specific, this assumes schooling has a positive relationship to skill. With this, one would normally expect a rise in years of schooling attributable to background characteristics to result in a rise in skill. Hence, the reverse may be the case, where a fall in years of schooling results in a loss in skill. However, from the output (Table 2.12), columns (1) and (2) suggest a loss of 2.5 years of schooling attributable to having a father with post-secondary schooling has no effects on skill. Hence, to draw a useful inference from this anomaly (all things being equal and without any inherent defects in methods), a consideration for the efficiency of schooling—the extent to which schooling raises skill—will give a useful understanding of this anomaly. Affirming schooling is efficient in Kenya would mean the premise—mediating effects of schooling in the relationship between parental education and skill—is well founded and not a mere assumption. This is examined in the subsequent subsection. As earlier highlighted, besides, this effect of the father's post-secondary schooling, the finding from the use of the indicator of the mother's post-secondary schooling as the basis of treatment (columns (3) and (4) of Table 2.12) is consistent with the father's post-secondary schooling. However, in the DiD approach, the socioeconomic statuses (high- and low-SES) have consistent effects on schooling and skill, unlike the use of parental post-secondary schooling that impacts schooling but not skill. Hence, the latter (parental post-secondary schooling) is the focus of this (subsequent) analysis. However, as usual, other measures of background characteristics such as socioeconomic status, are used to assess the robustness of the effects of parental education. Having discussed the need for further analysis of the efficiency of schooling in urban Kenya, I now discuss the approach to (further) test the validity of estimates of the DiD approach.

The Need for the 2SLS-IV Approach

So far, the evidence from the DiD estimand suggests both parental education and socioeconomic status have no effect on skill even with substantial evidence of the impact of the reform on the skills of the disadvantaged, particularly, for those with parents without post-secondary schooling. It is argued that background characteristics (parental education) impact skill through schooling which makes the effects of parental education on skill not merely dependent on the reform but (substantially) on the prevailing efficiency of schooling (discussed in the next subsection), the limitation of the ATET (please, see the empirical framework) that give causal estimates of the effects of parental education that relate to the treated only (and not the population of interest) may mean that the DiD approach may fail to reliably capture estimates of the causal effect of background characteristics on skills attributable to the reform in Kenya. This suggests the need for estimand that gives an ATE (Average Treatment Effect). However, the limitations to the scope and data of this study mean obtaining an ATE (see empirical framework) is a step to explore in future studies, however, the ATET obtained, provides a useful first step to drawing causal inference from estimates. Furthermore, in subsequent analysis to get closer to the ATE, I examine the efficiency of schooling with skill as an outcome (schooling as inputs) in Ordinary Least Squares estimation and extend this to the two-stage least squares instrumental variables (2SLS-IV) estimations to obtain Local Average Treatment Effects (LATE) of schooling on skill. The flexibility in this approach makes it possible to examine the reform-affected estimates (specifically, for those impacted by the reform) of the effects of schooling on the skill of most subcategories of interest in this analysis. Hence, this subsampling or treatment heterogeneity across measures of background characteristics gives useful insights into how parental education and socioeconomic status impact skill through their impacts on schooling. This brings us closer to the ATE as it strictly considers the ATE for those impacted by the reform and the subsampling makes it possible to give some useful consideration to the treated category. This is another way to test the robustness of ATET that considers the effects of the reform on the treated category. Hence the LATE will provide a further robustness check on the ATET.

Table 2-12 Effects of Background Characteristics on Schooling and Skill

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Father with Post-Sec Education		Mother with Post-Sec Education		Low SES at Age 15		High SES at Age 15	
	Schooling	Skill	Schooling	Skill	Schooling	Skill	Schooling	Skill
Reform Indicator	0.695**	0.189**	0.741**	0.243***	0.528	0.225**	0.960***	0.283***
	(0.005)	(0.002)	(0.002)	(0.000)	(0.064)	(0.001)	(0.000)	(0.000)
Indicator of Father's Post-Secondary Ed	5.435***	0.869***						
	(0.000)	(0.000)						
Reform × Father	-2.529**	-0.232						
	(0.008)	(0.105)						
Indicator of Mother's Post-Second Ed.			5.628***	0.735***				
			(0.000)	(0.000)				
Reform × Mother			-2.613*	-0.153				
			(0.019)	(0.386)				
Low SES at Age 15					-2.579***	-0.387***		
					(0.000)	(0.000)		
Reform × Low SES					0.721	0.062		
					(0.101)	(0.579)		
High SES at Age 15							2.660***	0.467**
							(0.001)	(0.002)
Reform × High SES							-1.005	-0.183
							(0.191)	(0.258)
Nairobi_Strata	0.809*	-0.250**	0.689	-0.267**	0.667	-0.282**	0.803	-0.259**
	(0.043)	(0.006)	(0.122)	(0.008)	(0.112)	(0.002)	(0.059)	(0.005)
Large_Strata	0.664	-0.226**	0.674	-0.225*	0.674	-0.230*	0.718	-0.222*
	(0.053)	(0.005)	(0.074)	(0.014)	(0.079)	(0.016)	(0.062)	(0.019)
Medium_Strata	0.761**	-0.241***	0.744*	-0.241**	0.828*	-0.229**	0.954**	-0.207*
	(0.010)	(0.000)	(0.029)	(0.002)	(0.028)	(0.006)	(0.009)	(0.010)
Small_Strata (Reference)								
_cons	9.572***	0.118	9.593***	0.122	10.46***	0.237**	9.306***	0.053
	(0.000)	(0.093)	(0.000)	(0.087)	(0.000)	(0.007)	(0.000)	(0.458)
N	3249	3249	3478	3478	3693	3693	3693	3693
R-sq	0.130	0.090	0.086	0.057	0.063	0.041	0.036	0.029
adj. R-sq	0.128	0.088	0.085	0.055	0.061	0.039	0.034	0.027

Note: Table reports outputs of variants of models 2.5; and 2.6 following the Difference-in-Differences specifications as columns (1), (3), (5), and (7); and (2), (4), (6) and (8) respectively. The outcome of the former is schooling as actual years of schooling and the latter skills, as standardised reading proficiency. Both models assess the effects of the 1985 curriculum structural reform, p1985; and background characteristics (as indicators of the father's tertiary education; mother's tertiary education; low socioeconomic status; and high socioeconomic status, with mid-level socioeconomic status as the reference category). Controls include strata-specific effects which is a basis of the stratified sampling (based on the number of households across cities) across Nairobi, strata_N, other large cities, strata_L with over 100 000HH, medium cities, strata_M with over 60 000HH but under 100 000HH, and other cities under, strata_S with under 60 000HH (reference category). The p-values in parentheses: * $p<0.05$; ** $p<0.01$; *** $p<0.001$. With clustered robust standard error, at the district level.

2.3.3 Effects of Schooling on Skill: 'Efficiency' Consideration

Evidence using the baseline models (for schooling and skill) and extensions such as the Difference-in-Differences specification suggests that background characteristics (substantially) explain skills through its mediation with schooling (comparing estimates in Tables 2.11 and 2.12). As argued, there is a need to examine the effects of schooling on skill to improve understanding of the effects of background characteristics on skill. In line with the objective of this study, I argue that the effects of background characteristics on skill, are at least, in part, dependent on the 'efficiency or productivity in schooling' as background characteristics are assumed to impact skill by their impacts on schooling in Kenya. In unravelling the effects of schooling on skills, beyond examining this directly by assessing the extent to which schooling explains skills, I consider the effects of inequality (and inequity) in schooling by examining the effects of schooling on skill across subsamples of based on background characteristics. To do this, I deploy the Oaxaca-Blinder Decomposition technique that decomposes estimates of differentials into 'characteristic' and 'return' components that give insights into effects attributable to inequality and inequity (or potential discrimination) respectively. I then turn to obtain estimates of the effects of schooling on skill from which causal inferences are drawn. I exploit the reform as an instrument (instrumenting schooling) in a quasi-experimental (Instrumental Variables) approach that uses the Two-Stage Least Squares (2SLS-IV) to draw causal inferences from estimates of the effects of schooling on skill. Examining relevant sub-samples, I obtain (causal) estimates of the effects of schooling on the skills of the treated (based on background characteristics). Assuming background characteristics 'substantially' impact skill by its impacts on schooling, I argue that IV estimates give the (causal) effect of background characteristics (and related reforms) on skills. This further supports findings on the effects of background on skill from the DiD analysis.

What is the Efficiency of Schooling in Urban Kenya?

Figure 2.6 presents some non-parametric evidence of the (irregular) variations in skill (reading proficiency) across credential categories. The kernel density estimate suggests schooling explains skills in urban Kenya, evidence of the irregular skill distribution persists for respondents across all the categories of education, in Kenya. This suggests evidence of other factors resulting in the anomaly (see motivation/introduction for a detailed discussion of the anomaly).

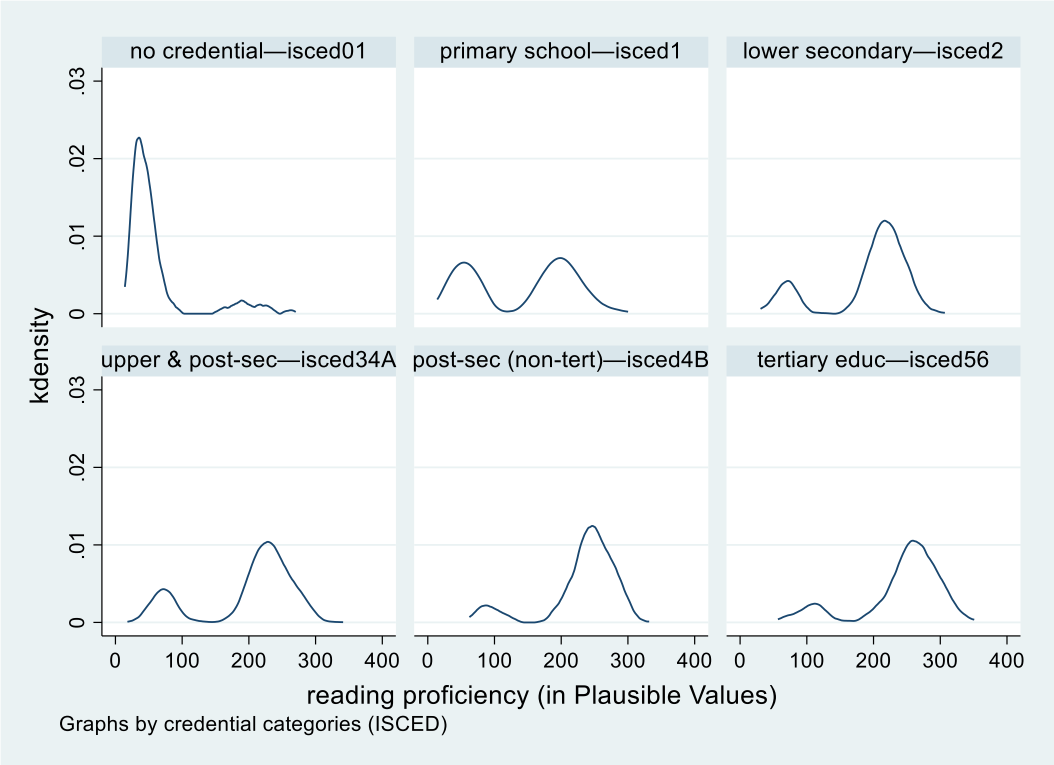


Figure 2-6 Kernel Density Plot, Reading Proficiency, by Educational Attainment.
Source: STEP Skills Measurement Program of the World Bank.

Table 2.13 presents the outputs of a variant of Equation 2.6. The model accounts for background characteristics, using the father's post-secondary schooling as a predictor of skill (as an outcome) only. The measure of low socioeconomic status is statistically insignificant hence it is excluded as other measures of background characteristics are substantially subsumed by the effect of the father's post-secondary schooling. This specification is based on the output of column (16) of Table 2.11. The model accounts for schooling (continuous and categorical) as useful predictors of interest. In addition to this, accounted for, are district-specific effect and average schooling at the district level which enters as a quadratic term, I alternate the latter with stratum characteristics as argued in the previous subsection, as accounting for both stratum characteristics and average schooling across districts shows some spurious effects. Also controlled for, are typical demographics such as age and gender. Outputs presented show specifications of the model (Equation 2.6) that have at least, an explanatory power of 36.6% of the outcome (skill or reading proficiency). Table 2.13 presents the outputs from estimating several specifications of the model (based on Equation 2.6). Although the focus is on the variables of interest in line with the core testable predictions (research questions—effects of schooling on skill), several other estimates or variables in the outputs are worth mentioning as a few of these[72] are pertinent to the objective of this research.

Table 2.13 suggests, regardless of the credential category, a one-year increase in schooling explains a standard rise of over thirteen (13) percentage points in cognitive skills, statistically significant at a 0.1% level (columns 1-6). This is robust to accounting for the indicator of father's post-secondary education, age, gender district-specific and strata-specific or average (or district-level) years of schooling across districts which enter as a quadratic term. Evidence from the quadratic effects of average or district-level schooling suggests, a negative (or an adverse) effect on adult skills. This may become positive with a substantial rise in the levels of schooling in districts. The statistically significant effect of average schooling in districts is a crucial understanding of this study that also examines externalities of schooling (please, see Chapter 4 devoted to examining externalities of schooling). Whilst accounting for the average schooling across districts is deemed to provide more plausible estimates of the effects of individual-level schooling on skill, evidence (see columns (5) and (6)) suggests, regardless of the model specification (accounting for strata-specific effects or average schooling in districts), the effect of schooling on skill remains the same.

[72] Across Table 2.13, ***evidence suggests, the size (or number of households) of districts, gender and age are negatively associated with cognitive skills.*** Hence, being in districts with under 60 000HH favourably impacts skills relative to those with over 60 000HH. Relative to the young, ageing is associated with lower cognitive skills. Similarly, relative to being a male, being a female is associated with low cognitive skills. These shortfalls in skills across age and gender categories may be attributable to the variations in the schooling of the old and the female relative to their young and male counterparts, respectively. This makes the former consistent with argument on the rise in educational attainment over time in sub-Sahara Africa (Barro and Lee, 2013) as the older generation have (substantially) fewer years of schooling relative to the younger generation. The latter is consistent with the study of Klugman et al., 2014 for sub-Sahara Africa and in developing contexts (see Alderman et al, 1998; Behrman and Knowles, 1999), where evidence suggests relative to the male gender, access to schooling for the female gender is substantially restricted. This is reflected in the substantial difference in the educational attainment of fathers relative to those of mothers which may further explain the differences in the schooling of their wards (across genders) in an intergenerational manner.

Table 2-13 OLS Estimates of the Effects of Schooling and Background on Skill

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	zapvlit_c (standardised reading proficiency as a measure of cognitive skill)											
father_educ_456_	0.222***	0.237***	0.228***	0.206***	0.215***	0.208***	0.229***	0.242***	0.230***	0.207***	0.213***	0.206***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Yos	0.133***	0.139***	0.138***	0.130***	0.136***	0.136***	0.006	0.020	0.013	0.002	0.015	0.0090
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.626)	(0.180)	(0.346)	(0.862)	(0.313)	(0.520)
sc_1 (Yos X isced1)							0.110***	0.103***	0.107***	0.110***	0.103***	0.106***
							(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
sc_2 (Yos X isced2)							0.140***	0.125***	0.133***	0.139***	0.124***	0.132***
							(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
sc_34A (Yos X isced34A)							0.120***	0.113***	0.117***	0.120***	0.113***	0.117***
							(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
sc_4B (Yos X isced4B)							0.124***	0.115***	0.120***	0.125***	0.117***	0.122***
							(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
sc_56 (Yos X isced56)							0.118***	0.108***	0.113***	0.119***	0.111***	0.116***
							(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age				-0.006**	-0.008***	-0.007**				-0.006**	-0.008***	-0.007**
				(0.007)	(0.000)	(0.002)				(0.008)	(0.001)	(0.003)
Gender				-0.098***	-0.067*	-0.083**				-0.082**	-0.058*	-0.071*
				(0.001)	(0.034)	(0.006)				(0.003)	(0.048)	(0.012)
stratum_N		-0.300***			-0.311***			-0.260***			-0.273***	
		(0.000)			(0.000)			(0.000)			(0.000)	
stratum_L		-0.259***			-0.277***			-0.216***			-0.235***	
		(0.000)			(0.000)			(0.001)			(0.000)	
stratum_M		-0.287***			-0.307***			-0.235***			-0.257***	
		(0.000)			(0.000)			(0.000)			(0.000)	
avg_yos_			-0.243**			-0.256**			-0.206**			-0.216**
			(0.003)			(0.003)			(0.005)			(0.006)
avg_yos_(squared)			0.010**			0.011**			0.009**			0.0095*
			(0.009)			(0.009)			(0.010)			(0.012)
_cons	-1.399***	-1.194***	0.025	-1.135***	-0.876***	0.381	-1.315***	-1.153***	-0.153	-1.052***	-0.832***	0.196
	(0.000)	(0.000)	(0.953)	(0.000)	(0.000)	(0.420)	(0.000)	(0.000)	(0.689)	(0.000)	(0.000)	(0.658)
N	3386	3386	3386	3386	3386	3386	3386	3386	3386	3386	3386	3386
R-sq	0.366	0.381	0.372	0.371	0.388	0.378	0.386	0.396	0.389	0.391	0.402	0.395
adj. R-sq	0.365	0.380	0.371	0.371	0.387	0.377	0.385	0.394	0.388	0.389	0.400	0.393

Note: Table reports outputs of variants of Equation 2.6, as columns 1-6; and 7-12, respectively. With skill as standardised reading proficiency as the outcome. The predictors of interest are the continuous measure of schooling, years_educ_act and the product term of continuous and categorical schooling (please, see the empirical framework for details of this product term). Controls include strata-specific effects (based on the number of households across cities) and include, Nairobi, strata_N, other large cities, strata_L with over 100 000HH, medium cities, strata_M with over 60 000HH but under 100 000HH, and other cities under, strata_S with under 60 000HH (reference category). Other control variables include the average schooling, avg_yos, across districts, which also enters as a quadratic term. Other covariates include age and an indicator of female gender. p-values in parentheses: * $p<0.05$; ** $p<0.01$; *** $p<0.001$. With clustered robust standard error, at the district level.

Columns (7)-(12) report estimates of the effects of schooling on skill, accounting for credential categories. This is inspired by the study of Lee and Wie (2017) and gives useful insights into the efficiency of schooling or the extent to which schooling raises skill. One would expect (relatively) high credentials to explain (relatively) high reading proficiency, however, this is not the case in urban Kenya. The evidence suggests, relative to other credential categories, an additional year spent in lower secondary education (isced2, as sc_2) best explains reading proficiency, in urban Kenya. This is the case across columns (7)-(12) where with a reference to the years_educ_act variable that captures the effect of the schooling of those with no credentials (the reference category), the estimates of the joint effects of the categorical and continuous schooling (sc_1, sc_2, sc_34A, sc_4B and sc_56) present interesting findings on the efficiency of schooling in urban Kenya. The effects of years_educ_act across (7)-(12) is statistically insignificant. Hence, for those with no formal qualification, an additional year of their limited schooling (if any) does not raise skill or reading proficiency. Specifically, relative to this reference category, having lower-secondary education, an additional year of schooling is associated with a standard rise of 13.2 percentage points in reading proficiency. However, an additional year of schooling at higher (or tertiary) education does not necessarily result in higher returns (in skills) or reading proficiency.

The evidence suggests that having each of the isced34 and isced4B credentials, an additional year of schooling is associated with a standard rise of 11.7 and 12.2 percentage points in reading proficiency respectively. For tertiary schooling (isced56), the effect is only 11.6 percentage points in cognitive skills. Although effects are statistically significant at the 0.1% level, the fall in the effects of an additional year of schooling, from lower-secondary to higher education, is attributable to the inefficiency in schooling across isced34A to isced56, relative to isced2 category of schooling (see column (12) of Table 2.13). So far, evidence of the effects of schooling on skill is generally positive and statistically significant in urban Kenya. Hence, suggesting schooling (substantially) explains skill. However, further evidence suggests that the substantial variation in the effects of an additional year of schooling (across credential categories) on adult skills is attributable to inefficiency in schooling which gives insights into issues of quality in the provision of schooling, in urban Kenya.

Although previous analysis accounts for father’s post-secondary schooling as a measure of background characteristics, to further examine how inequality (or inequity) in schooling attributable to background characteristics impacts skill, I consider heterogeneity across subsamples based on background characteristics, deploying the Oaxaca-Blinder decomposition methods that aid a thorough analysis of inequality/inequity in schooling impacts skill or reading proficiency, in urban Kenya.

To do this, I use a similar specification to column 12 on Table 2.13, accounting for demographics such as age, gender, district-specific effects, and average schooling across districts. The output of the first-stage model which is further decomposed by characteristics, return, and interaction effects using the Oaxaca-Blinder Decomposition methodology is given as thus:

Table 2-14 Effects of Schooling on Skill, by Father's Education

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable: Cognitive Skill (PV of Reading Proficiency) _standardised					
	Pooled		Father's Post-Secondary Education (indicator)			
			1 ('advantaged' or 'treated')		0 ('disadvantaged' or untreated)	
Years_of_Schooling	0.142***	0.027	0.102***	-0.037	0.141***	0.008
	(0.000)	(0.078)	(0.000)	(0.350)	(0.000)	(0.556)
Years_of_Schooling X ISCED1		0.088***		0.166***		0.107***
		(0.000)		(0.000)		(0.000)
Years_of_Schooling X ISCED2		0.116***		0.135***		0.134***
		(0.000)		(0.000)		(0.000)
Years_of_Schooling X ISCED34A		0.102***		0.142***		0.120***
		(0.000)		(0.000)		(0.000)
Years_of_Schooling X ISCED4B		0.109***		0.146***		0.127***
		(0.000)		(0.000)		(0.000)
Years_of_Schooling X ISCED56		0.105***		0.142***		0.122***
		(0.000)		(0.000)		(0.000)
Age	-0.008***	-0.009***	-0.012**	-0.012**	-0.008***	-0.008***
	(0.000)	(0.000)	(0.003)	(0.003)	(0.000)	(0.000)
Gender	-0.079**	-0.061*	-0.056	-0.052	-0.080**	-0.066*
	(0.008)	(0.028)	(0.260)	(0.291)	(0.009)	(0.023)
Average Years of Schooling	-0.329***	-0.272***	0.153	0.086	-0.287***	-0.240**
	(0.000)	(0.000)	(0.262)	(0.478)	(0.000)	(0.001)
Av. Years of Schooling (Squared)	0.015***	0.012***	-0.005	-0.002	0.013**	0.011**
	(0.000)	(0.001)	(0.405)	(0.657)	(0.001)	(0.002)
_cons	0.711	0.502	-1.543	-1.166	0.512	0.339
	(0.119)	(0.244)	(0.056)	(0.108)	(0.256)	(0.420)
N	3875	3875	679	679	3386	3386
R-sq	0.376	0.392	0.200	0.210	0.372	0.389
adj. R-sq	0.375	0.390	0.193	0.197	0.371	0.387

Note: Table reports the First Stage of the Oaxaca-Blinder. This shows the effects of schooling and demographics based on the father's post-secondary education. p-values in parentheses: * $p<0.05$; ** $p<0.01$; *** $p<0.001$, with robust standard error, clustered at the district level.

A disaggregation based on background characteristics (here, based on the indicator of the father's post-secondary schooling) unravels interesting insights. To start with, comparing the effects of continuous schooling on skill (see (1), (3), and (5) of Table 2.13), the evidence suggests that, regardless of the father's education and credential categories (using continuous schooling), an additional year of schooling explains a standard rise of 14.2% points in cognitive skill (column (1)). However, disaggregating this based on the father's post-secondary schooling, the evidence suggests that, for those who have fathers with no post-secondary education (the ‘disadvantaged’), an additional year of schooling explains a standard rise of 14.1 percentage points in skill (column (5)) but then having a father with post-secondary education (the ‘advantaged’), an additional year of schooling explains a standard rise of only 10.2 percentage points in skill (column (3)). All estimates are statistically significant at the 0.1% level. Interestingly, this difference in the effect of schooling on skill based on the father’s post-secondary schooling is consistent with previous findings on the effects of background characteristics on schooling (Table 2.12) where the DiD estimand shows a reform-affected rise in years of schooling attributable to having a father without post-secondary schooling and reform-affected fall in years of schooling attributable to having a father with post-secondary schooling. It is important to note that a rise/fall in years of schooling is not the same as overall educational attainment.

The evidence from Table 2.11 suggests that those who have fathers with post-secondary schooling are more educated. Hence, the ‘advantaged’ have attained higher schooling, relative to the ‘disadvantaged’. Suggesting the higher skill for most of those who have fathers with no post-secondary schooling, relative to those who have fathers with post-secondary education as reported in columns (5) and (3) of Table 2.14 which is supported by the earlier findings on the inefficiency of schooling in urban Kenya, where years of schooling for those who have attained less schooling (isced2) explains substantial skill or reading proficiency, relative to the effects of years of those that have attained more schooling.

Interestingly, disaggregating (across credential categories) schooling attributable to the father’s tertiary education, see columns (2), (4), and (6). Evidence suggests some variations in the effects of schooling (accounting for credential categories) on skills based on background characteristics (here, using the father’s post-secondary schooling as the basis of treatment). The evidence suggests those who have fathers with no post-secondary education (column (6)) have similar trends consistent with the pool (column (2)) as previously discussed where the peak of returns is at isced2 which falls and rises slightly. However, this is not exactly the case for those who have fathers with post-secondary schooling (column (4)) where the findings suggest that those who have fathers with post-secondary education may have received more inefficient schooling as returns to (or skills from) an additional year of their schooling is greatest at the least category of schooling (isced1) and falls drastically afterward, rising slightly again.

Put together, a substantial variation in the efficiency of schooling received on this basis of the father's post-secondary education, in that, the category of schooling that yields the least returns for those that have fathers with no post-secondary education (isced1) is the category of schooling that yields the greatest returns for those that have fathers with post-secondary education. Interestingly, the credential category that yields the least returns for respondents having fathers with post-secondary education (isced2) is the same as the credential category that yields the greatest returns for respondents with fathers with no post-secondary education. However, it is interesting that across each of the categories of schooling, those that have fathers with post-secondary education and those that have fathers with no post-secondary education (comparing effects of categorical schooling in columns (4) and (6)), evidence suggests higher returns for an additional year of schooling for those that have fathers with post-secondary education relative to same schooling categories of those that have fathers with no post-secondary education. Hence, findings for 'categorical schooling' contrast findings for 'continuous schooling' across the advantaged (respondents having fathers with post-secondary education) and the disadvantaged (respondents having fathers with no post-secondary education).

Compared to the advantaged, the model fit for the disadvantaged shows schooling, gender, age, district-specific effect, and the average schooling across districts explain about 39% of the variation in skills, however, this is about 21% for the advantaged showing several factors (unaccounted for) may explain the skills of the advantaged, this may include nature, genetic and factors that may be potentially discriminating (subject to a further analysis). As further checks on the robustness of estimates based on the father's post-secondary schooling, the evidence suggests (see Table 2.16) that, although using the indicator of mother's post-secondary education as the basis of treatment shows some consistency. Particularly, relative to respondents who have mothers with post-secondary education, respondents who have mothers without post-secondary education have a higher return to an additional year of schooling without accounting for credential categories. However, accounting for the effects of an additional year of schooling across credential categories suggests that, only respondents who have mothers without post-secondary education have statistically significant effects of their schooling on their skill, however, the trend shows evidence of inefficient schooling for these.

Somewhat consistent with the findings from the DiD analysis, the evidence suggests no substantial variation in the effects of an additional year of schooling (without accounting for credential categories) on skill based on socioeconomic status. However, accounting for credential categories, the evidence suggests substantial variations in the effects of schooling on skill based on socioeconomic status, which is in line with the effect of schooling on skill using the father's post-secondary education as a basis for the 'treated'. The evidence suggests that, relative to those with low socioeconomic status, those with high socioeconomic status have higher returns to an additional year of schooling and the evidence of inefficient schooling persists regardless of socioeconomic status, as those with low levels of schooling have higher returns to an additional year of their schooling, relative to those that attained higher levels of schooling.

I now turn to present a summary of key findings so far. Without accounting for credential categories or background characteristics, an additional year of schooling (using continuous schooling) significantly impacts cognitive skills in urban Kenya. However, disaggregating this across credential categories suggests that schooling in urban Kenya is inefficient, as overall, those that have attained relatively low levels of schooling (isced1 for the 'advantaged'; and isced2 for the 'disadvantaged') have substantially high reading proficiency compared to those that have attained relatively high credential categories. The consideration for background characteristics unravels interesting findings. The evidence suggests that on average (using continuous schooling), the return to an additional year of schooling for the 'advantaged' is high relative to the 'disadvantaged'. However, across all schooling categories, the return to an additional year of schooling is greater for the 'advantaged', relative to the 'disadvantaged'. Although it is concluded that the inefficiency in schooling—that suggests those with low credential category have higher returns to their schooling or higher skills from their schooling—coupled with the relatively low educational attainment for those that have fathers with no post-secondary schooling make it possible for those that have fathers with no post-secondary schooling (relative to those have fathers with post-secondary schooling) to have higher returns to an additional year of schooling (without accounting for credential categories). In addition to this, another contradictory point that may have resulted in the statistically insignificant effects of the DiD estimand (the effects of background characteristics on skill) is that relative to the 'disadvantaged', the 'advantaged' have higher returns to an additional year of schooling, across all credential categories. It is important to note that the DiD estimand provides the Average Treatment Effects on the Treated (ATET). Hence, the 2SLS-IV estimation that gives the Local Average Treatment Effect (LATE) aiding useful comparisons based on background characteristics, may provide more useful evidence that brings us closer to drawing causal inferences on the effects of background characteristics on skill. Although ideal (as the Average Treatment Effect (ATE)). However, before the 2SLS-IV analysis, I further examine skill, specifically, by assessing the absolute skill gap between the 'advantaged' and the 'disadvantaged', particularly, how inequality or inequity in schooling impacts the skill gap in urban Kenya, to fully unravel the evidence presented in the nonparametric analysis (see Kernel density plots, Figures 2.6 and 2.7).

Explaining the Skill Differential: The Oaxaca-Blinder Decomposition

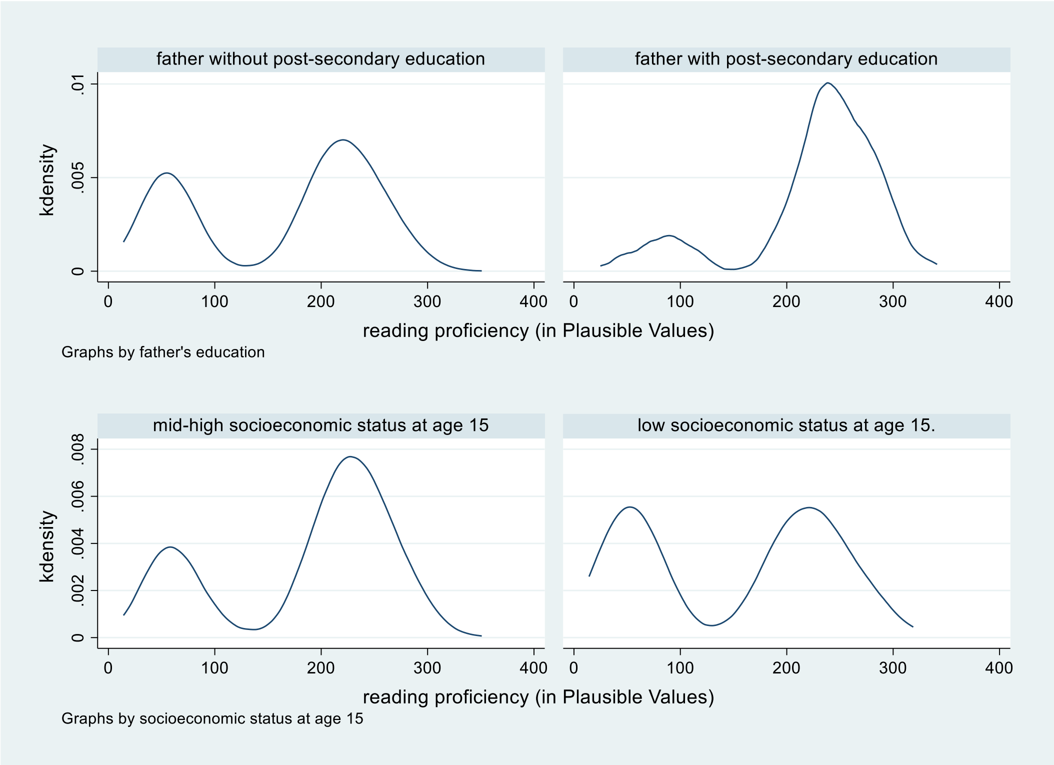


Figure 2-7 Kernel Density Plot, Reading Proficiency, by Backgrounds.
Source: STEP Skills Measurement Program of the World Bank.

I now discuss the findings of the Oaxaca-Blinder Decomposition which gives useful insights into the inequality in skill acquisition, emphasising access to schooling attributable to background characteristics. Accounting for schooling, age, gender, district-specific (fixed) effects, and the average (or district-level) years of schooling, which enters as a quadratic term. Table 2.15 summarises the estimates of the skill gap between those who have fathers with post-secondary education and those who have fathers without post-secondary education (based on Table 2.14). In addition to this, reported in Table 2.15 is the decomposition of the skill gap into the characteristic, return, and interaction effects are reported in Table 2.15.

Table 2.15 shows that having a father with post-secondary education is associated with a mean standard rise of 59.8 percentage points in cognitive skills, whereas having a father with no post-secondary education is associated with a standard fall of 8.76 percentage points, in cognitive skills.

Table 2-15 Oaxaca-Blinder Decomposition, by Father's Education

Differential								
Prediction 1 (Having a father without Post-Secondary Education)					-0.088**			
					(0.002)			
Prediction 2 (Having a father with Post-Secondary Education)					0.598***			
					(0.000)			
Difference					-0.686***			
					(0.000)			
	Total	YoS	Age	Gender	Avg_Yos	Avg_Yos (Squared)	District (fixed) Effect	Constant
Characteristic/Endowment	-0.443***	-0.355***	-0.048**	-0.0003	-0.125	0.084	-0.0004	
	(0.000)	(0.000)	(0.003)	(0.804)	(0.272)	(0.410)	(0.820)	
Return Effect/Coefficients	-0.204***	0.505**	0.140	-0.016	-5.015*	2.166	-0.134	2.151*
	(0.000)	(0.006)	(0.213)	(0.609)	(0.027)	(0.076)	(0.209)	(0.044)
Interaction/Joint Effects	-0.039	-0.132**	0.022	-0.0002	0.372*	-0.302	0.002	
	(0.413)	(0.006)	(0.216)	(0.820)	(0.050)	(0.104)	(0.744)	

Note: This Table reports the Second Stage of the Oaxaca-Blinder Decomposition, showing the effects of schooling and demographics based on the father's post-secondary education. p-values in parentheses: * p<0.05; ** p<0.01; *** p<0.001, with robust standard error, clustered at the district level.

Between respondents having fathers with post-secondary education (also referred to as the 'advantaged') and the respondents having fathers with no post-secondary education (also referred to as the 'disadvantaged'), the evidence suggests a skill gap of a standard 68.6 percentage points in cognitive skills. Based on the specified model (see empirical strategy), for the advantaged, having the characteristics (or endowment) of the disadvantaged explains a mean standard fall of 44.3 percentage points in cognitive skills. A standard fall of 20.4 percentage points in cognitive skills for the advantaged is explained by the return (or coefficient) effect, of the disadvantaged. Hence, 65% (44.3/68.6) of the skills gap between the 'advantaged' and 'disadvantaged' is attributable to the effects of inequality defined by differences in characteristics (or endowment), roughly 30% (20.4/68.6) of the gap in skills between these is attributable to the effects of potential discrimination or inequity defined by difference in the returns (coefficient). Further decomposition of the skill gap suggests evidence of the inequality and inequity (potential discrimination) in skill formation are mainly driven by differences in characteristics and returns to individual and average (district-level) schooling between the 'advantaged' and 'disadvantaged'.

Interestingly, evidence from Table 2.15 shows that although the overall (see column titled Total) effects of the Returns and Characteristics show evidence of substantially high effects of returns and characteristics of the advantaged over the disadvantaged as drivers of the skill gap. The positive effect of returns to individual schooling is consistent with the results from previous subsections—which show higher returns to education of the disadvantaged (without accounting for credential categories), relative to the advantaged—further findings from Table 2.15 confirm the higher returns to the schooling of the disadvantaged is at least, in part, due to the inefficiency in schooling, in urban Kenya and the relatively low educational attainment of the disadvantaged (evident from the relatively high effect of the schooling characteristic of the advantaged), in urban Kenya. Hence, this evidence further suggests that the inequity attributable to the return effects (and possible interaction of return and characteristic effects) may substantially explain the skill gap between the advantaged and the disadvantaged. Further evidence suggests, this is not attributable to the return effects of individual schooling; or the interaction of the characteristics and return effects of schooling (at the district level) as these favourably impact the skill of the disadvantaged relative to the advantaged. However, the effect of the interaction of characteristics and return to individual schooling; and the mean effects of average schooling (at the district level)[73] are two factors in addition to other unobserved factors (see the statistically significant constant/intercept term) that may be driving inequity in skill in urban Kenya.

Using the indicator of the mother's post-secondary education, similar outcomes in using the indicator of the father's post-secondary education are evident (see Table 2.17) in decomposing the same models. Specifically, the outcomes indicate similar skill gaps, characteristics, and returns between the advantaged and the disadvantaged. However, more substantial mean and statistically significant effects are found using the indicator of the father's education. Using indicators of socioeconomic status as the indicator of background characteristics, no statistically significant effect is attributable to potential discrimination.

Besides improving understanding of the drivers of inequity and inequality in skills, the Oaxaca-Blinder decomposition gives further insights — improving understanding of the inefficiency in schooling — and affirming arguments that, although the disadvantaged have higher returns to an additional year of their schooling (without accounting for credential categories), the substantial difference in the number of years of schooling (or educational attainment) between the advantaged and the disadvantaged best explains their skill gap. This suggests that the skill gap between the disadvantaged and the advantaged is through the effects of the differences in characteristics, rather than return effects. Hence, with the inefficiency in schooling (where schooling at low levels better raises skill), schooling

[73] A useful example of this case includes, having a father with tertiary education makes respondents be in better districts with (relatively) more efficient schooling or it may make it possible for such respondents to have higher parental engagement on their schooling which favourably impact schooling and skill. Alternatively, having a father with post-secondary schooling may make it possible to live and work in urban districts with good employment that further raise skill. Such, make it unfair for those that have fathers without tertiary education, hence, an issue of inequity based on father's education (as a background characteristic).

impacts skill by increase in schooling (effects of inequality or characteristic) rather than through potential discrimination (effects of inequity or returns). Hence, whilst the gap in skills of the ‘advantaged’ over the ‘disadvantaged’ come through substantial effects of inequity in skill (particularly, district-level schooling), the impact of individual schooling is substantially through its characteristic effect. Hence, overall, raising mere years of schooling can effectively raise skills in urban Kenya, amidst, the stark inefficiency in schooling.

Table 2-16 Effects of Schooling on Skill, by Background Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Dependent Variable: Cognitive Skill (PV of Reading Proficiency) _ standardised													
	Pool		Having a mother with Post-Secondary Education				With Low SES at Age 15				With High SES at Age 15			
			1		0		1		0		1		0	
Years_of_Schooling (YoS)	0.143***	0.026	0.102***	0.093	0.141***	0.024	0.140***	0.035	0.140***	0.022	0.138***	-0.011	0.142***	0.030
	(0.000)	(0.094)	(0.000)	(0.797)	(0.000)	(0.140)	(0.000)	(0.103)	(0.000)	(0.198)	(0.000)	(0.698)	(0.000)	(0.058)
YoS X ISCED1		0.088***		0.073		0.088***		0.088***		0.086***		0.138***		0.085***
		(0.000)		(0.834)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)
YoS X ISCED2		0.117***		0.034		0.117***		0.104***		0.119***		0.167***		0.112***
		(0.000)		(0.923)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)
YoS X ISCED34A		0.104***		0.026		0.103***		0.097***		0.104***		0.143***		0.099***
		(0.000)		(0.942)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)
YoS X ISCED4B		0.111***		0.038		0.109***		0.103***		0.110***		0.148***		0.106***
		(0.000)		(0.915)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)
YoS X ISCED56		0.107***		0.025		0.108***		0.101***		0.107***		0.141***		0.103***
		(0.000)		(0.943)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)
Age	-0.008***	-0.008***	-0.016**	-0.017**	-0.008***	-0.008***	-0.010***	-0.010***	-0.008***	-0.008**	-0.007	-0.007	-0.008***	-0.009***
	(0.000)	(0.000)	(0.002)	(0.002)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.053)	(0.104)	(0.000)	(0.000)
Gender			-0.139	-0.149*	-0.082**	-0.063*	-0.089	-0.058	-0.074*	-0.060*	-0.064	-0.046	-0.087**	-0.068*
			(0.062)	(0.050)	(0.010)	(0.038)	(0.213)	(0.407)	(0.016)	(0.038)	(0.403)	(0.536)	(0.004)	(0.014)
Average YoS	-0.338***	-0.277***	-0.001	-0.001	-0.002*	-0.002*	-0.005***	-0.005**	-0.001	-0.001	0.001	0.001	-0.002*	-0.002*
	(0.000)	(0.000)	(0.815)	(0.756)	(0.022)	(0.025)	(0.001)	(0.001)	(0.487)	(0.595)	(0.678)	(0.614)	(0.032)	(0.036)
Average YoS (Squared)	0.015***	0.013***	0.026	-0.072	-0.337***	-0.284***	-0.268	-0.238	-0.349***	-0.278***	-0.197	-0.139	-0.330***	-0.273***
	(0.000)	(0.001)	(0.883)	(0.681)	(0.000)	(0.001)	(0.064)	(0.072)	(0.000)	(0.000)	(0.143)	(0.234)	(0.000)	(0.000)
_cons	0.709	0.498	-0.667	-0.288	0.782	0.617	0.563	0.475	0.765	0.494	-0.067	-0.478	0.732	0.530
	(0.129)	(0.255)	(0.514)	(0.770)	(0.105)	(0.163)	(0.447)	(0.485)	(0.077)	(0.252)	(0.929)	(0.517)	(0.109)	(0.217)
N	3875	3875	414	414	3199	3199	965	965	2910	2910	423	423	3424	3424
R-sq	0.374	0.391	0.198	0.211	0.363	0.380	0.412	0.428	0.345	0.361	0.401	0.421	0.370	0.385
adj. R-sq	0.374	0.389	0.186	0.190	0.362	0.378	0.408	0.422	0.343	0.358	0.392	0.405	0.369	0.383

Note: This Table reports the First Stage of the Oaxaca-Blinder, showing effects of schooling and demographics based on the indicator of mother's post-secondary education; low socioeconomic status (Low SES); high socioeconomic status (High SES). p-values in parentheses: * $p<0.05$; ** $p<0.01$; *** $p<0.001$, with robust standard error, clustered at the district level.

Table 2-17 Oaxaca-Blinder Decomposition, by Mother's Post-Secondary Education

		Differential					
		Prediction 1 (Without Tertiary Qualified Mother)	-0.045				
			(0.136)				
		Prediction 2 (With Tertiary Qualified Mother)	0.591***				
			(0.000)				
		Difference	-0.635***				
			(0.000)				
	Total	Years_of_Schooling	Age	Gender	Avg_Yos	Avg_Yos ^2	Constant
Characteristic Effect/Endowment	-0.450***	-0.358***	-0.058**	-0.004	-0.018	-0.012	
	(0.000)	(0.000)	(0.002)	(0.298)	(0.881)	(0.909)	
Return Effect/Coefficients	-0.129**	0.522	0.212	0.029	-3.995	1.731	1.449
	(0.004)	(0.082)	(0.131)	(0.458)	(0.056)	(0.091)	(0.170)
Interaction/Joint Effects	-0.057	-0.135	0.030	0.002	0.257	-0.210	
	(0.446)	(0.083)	(0.136)	(0.524)	(0.083)	(0.122)	

Note: This Table reports the Second Stage of the Oaxaca-Blinder Decomposition, showing effects of schooling and demographics based on the indicator of mother’s post-secondary education. p-values in parentheses: * p<0.05; ** p<0.01; *** p<0.001, with robust standard error, clustered at the district level.

Table 2-18 Oaxaca-Blinder Decomposition, by Low Socioeconomic Status

		Differential					
		Prediction 1 (Without Low SES at age 15)	0.093**				
			(0.003)				
		Prediction 2 (With Low SES at age 15)	-0.270***				
			(0.000)				
		Difference	0.363***				
			(0.000)				
	Total	Years_of_Schooling	Age	Gender	Avg_Yos	Avg_Yos^2	Constant
Characteristic Effect/Endowment	0.317***	0.306***	0.011*	-0.0002	-0.043	0.039	
	(0.000)	(0.000)	(0.019)	(0.929)	(0.154)	(0.193)	
Return Effect/Coefficients	0.048	0.016	0.056	0.005	-0.704	0.338	0.127
	(0.137)	(0.866)	(0.512)	(0.906)	(0.574)	(0.630)	(0.836)
Interaction/Joint Effects	-0.002	0.004	-0.002	0.00002	-0.004	-0.011	
	(0.925)	(0.866)	(0.520)	(0.944)	(0.446)	(0.586)	

Note: This Table reports the Second Stage of the Oaxaca-Blinder Decomposition, showing effects of schooling and demographics based on the indicator of Low Socioeconomic Status. p-values in parentheses: * p<0.05; ** p<0.01; *** p<0.001, with robust standard error, clustered at the district level.

Table 2-19 Oaxaca-Blinder Decomposition, by High Socioeconomic Status

		Differential					
		Prediction 1 (Without High SES at age 15)	-0.032				
			(0.293)				
		Prediction 2 (With High SES at age 15)	0.270***				
			(0.000)				
		Difference	-0.302***				
			(0.000)				
	Total	Years_of_Schooling	Age	Gender	Avg_Yos	Avg_Yos^2	Constant
Characteristic Effect/Endowment	-0.225***	-0.219***	0.001	-0.002	0.010	-0.017	
	(0.000)	(0.000)	(0.820)	(0.534)	(0.682)	(0.514)	
Return Effect/Coefficients	-0.059	0.045	-0.032	-0.012	-1.386	0.660	0.799
	(0.153)	(0.677)	(0.750)	(0.749)	(0.254)	(0.264)	(0.216)
Interaction/Joint Effects	-0.018	-0.006	0.0001	-0.001	0.007	-0.011	
	(0.260)	(0.677)	(0.853)	(0.763)	(0.689)	(0.542)	

Note: This Table reports the Second Stage of the Oaxaca-Blinder Decomposition, showing the effects of schooling and demographics based on the indicator of High Socioeconomic Status. p-values in parentheses: * p<0.05; ** p<0.01; *** p<0.001, with robust standard error, clustered at the district level.

Making arguments on the efficiency of schooling; and inequality (or inequity) in skills attributable to the effects of background characteristics on schooling will be of policy relevance when causal inferences can be drawn from estimates. This is in line with the set objective to provide estimates from which causal inferences on the effects of schooling on skills (research question 2); and the effects of background characteristics on schooling and skill (research question 1).

In this subsection, the overarching finding is a response to research question 2, on the causal effect of schooling on skill. Further analysis involving subsampling (or treatment heterogeneity) from this helps to draw inferences in support of research question 1, on the effects of background characteristics on skill. In the previous analysis (following the DiD approach), the DiD estimand suggests parental education and socioeconomic status have no causal effect on skill. The limitations of the DiD estimand that gives the ATET at best, together with the inference that suggests, the effects of background characteristics on skill are through their dependence on the schooling (or efficiency of schooling) are argued to be reasons for this. These suggest the need for a more robust approach to determine the effects of background characteristics on skill. I now discuss the Two-Stage Least Square (2SLS-IV) approach.

2.3.4 The Effects of Schooling on Skill (LATE)

I now explore the Two-Stage Least Squares—Instrumental Variables (2SLS-IV) approach. I implement a 2SLS-IV with standardised skill as an outcome, instrumenting the individual and average schooling across districts with the interaction of the policy-dummy and quarter-of-birth indicators following the approach of Acemoglu and Angrist (1999) on the pecuniary returns to education. The average schooling across districts is accounted for in line with the study of Moretti (2006) who argues doing so improves estimates of the effect of schooling on earnings. These ideas are extended in this study, particularly, in this analysis that entails estimating the effects of schooling on skill. The effect of average (or district-level) schooling is discussed extensively as an externality of schooling in a subsequent chapter. In this chapter, the focus is on individual schooling. I implement the 2SLS-IV strategy in line with the objective of this study, firstly, I implement the skill production model using the full sample of the data, from this, I draw causal inferences on the average effects of schooling on skills, regardless of background characteristics. Secondly, I turn to fully account for each of the background characteristics by disaggregating the full sample based on each of the background characteristics, from this, I draw further inferences on the effect of background characteristics on the (causal) effects of schooling on skills. Whilst doing this does not give directly, the causal effects of background characteristics on skill, this, in comparison with previous analysis brings us closer to understanding the causal effects of background characteristics on skill.

Table 2.21 presents the outputs of the first stage 2SLS-IV model. Table 2.20 presents the output from the 2SLS-IV strategy. In this study, the 2SLS-IV strategy follows key ideas of the mainstream literature on the effect of schooling on earnings (see Card,1999). Inspired by the study of Acemoglu and Angrist (1999) that used the interaction of quarter of birth and year of birth; and compulsory attendance and child labour laws to instrument schooling, I use the interaction of quarter of birth and the reform dummy (see Data Chapter) as instruments (please, see further descriptions and tests of these instruments in the next chapter). In assessing the validity (relevance and exclusion restriction) of the reform as an instrument, whilst the reform shows useful evidence of (exogenous) variation in schooling, further evidence suggests the reform results in (exogenous) variation in skill as well (see evidence in Tables 2.11). This may suggest the exclusion restriction condition is not met, in the skill production function. In column 12 of Table 2.11, accounting for schooling, the nil effect of the reform suggests that the impact of the reform on skill is through its effects on schooling. This is supported by further inferences drawn from the DiD analysis, and the Oaxaca-Blinder decomposition, where the effect of schooling on skill suggests substantial dependence on the reform, the efficiency of schooling and background characteristics. I now turn to discuss the findings of the 2SLS-IV estimations.

Table 2-20 Effects of Schooling on Skill, the 2SLS-IV Approach

	Pool	Indic of Father's Post-Secondary Education		Indic of Mother's Post-Secondary Education		Indicator of Socioeconomic Status	
	(1)	(2)—High	(3)—Low	(4)—High	(5)—Low	(6)—High	(7)—Low
years_edu (yos)	0.318***	0.022	0.233*	0.006	0.207*	0.201	0.202***
	(0.000)	(0.734)	(0.027)	(0.931)	(0.029)	(0.085)	(0.000)
avg_yos (dist)	-0.245*	0.149	-0.243*	0.193	-0.217*	-0.124	-0.090
	(0.014)	(0.525)	(0.019)	(0.411)	(0.016)	(0.322)	(0.568)
_cons	-0.876	-1.358	0.019	-1.652	0.059	-0.838	-1.197
	(0.217)	(0.551)	(0.981)	(0.443)	(0.952)	(0.449)	(0.407)
N	3061	646	689	375	1109	351	697
R-sq	.	0.057	0.127	0.032	0.156	0.109	0.254
adj. R-sq	.	0.054	0.124	0.026	0.155	0.104	0.252

Note: Table reports outputs of the 2SLS-IV specifications that instruments individual and average schooling across districts using the interaction of quarter of birth and the reform indicator. Please, see Table 2.15.1 for the first-stage equation. Column (1) refers to the pool, columns (2) and (3) are subsamples based on the father's education; columns (4) and (5) are subsamples based on the mother's education; columns (6) and (7) are subsamples based on socioeconomic status. The outcome variable is skill, as standardised reading proficiency and the predictor of interest is schooling as actual years of schooling. No variables are controlled for in the model. The p-values in parentheses: * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

Table 2.20 presents outputs of the effects of schooling on skills on which causal inferences are drawn. This chapter focuses on the impacts of (individual) schooling on skill. Although controlled for, is the external returns (or externality) of schooling across districts, this is discussed extensively in the subsequent chapter. Comparing this IV to the OLS estimates (see Table 2.14 and 2.16) without accounting for background characteristics. The IV return estimate is at least two times the mean coefficients of similar OLS estimates, this is consistent with the mainstream literature that reports higher IV estimates (see Card, 1999). The IV and OLS estimates are statistically significant at the 0.1% level. Whilst this latter provides a useful estimate of the (causal) effects of schooling on skill, limitations in the modelling approach mean it may not be possible to account for returns to an additional year of schooling across credential categories. Hence, at best, this approach provides insights that support drawing causal inferences from estimates. This gives further insights on the claims made on the efficiency of schooling, in urban Kenya. I argue that the findings are consistent with the Local Average Treatment Effects (LATE), hence provide useful estimates from which causal inference on the effects of an additional year of schooling on cognitive skills is drawn.

Table 2.21 shows the outcome of the first stage of the 2SLS-IV, showing (exogenous) variation in the individual and district-level schooling attributable to the instruments. The evidence suggests substantial variation in individual and district-level schooling is attributable to the reform, particularly, in quarter 4. Evidence from the main analytical sample (pool) suggests that the reform raised individual (and average) schooling across districts by 1.925 (and 1.425) years of schooling respectively, statistically significant at the 0.1% level. These exogenous variations in schooling across the interaction of quarters of birth and the reform are deemed crucial drivers of the causal inference drawn from the substantial effects of individual schooling on skill (although without accounting for effects across credential categories). This results in a standard rise of 31.8 percentage points in reading proficiency for individuals; and a standard fall of 24.5 percentage points in reading proficiency (this average skill across districts is discussed in a subsequent chapter), please, see Table 2.16 columns (1) for this. However, in disaggregating this across background characteristics further insights are unravelled. The evidence from Table 2.21 shows, being born in quarter 2 to a father or mother with post-secondary education, explains a loss of about five (5) and two (2) years of schooling for individuals and at the district levels, respectively at the least, at 1% and 5% levels (see columns (3)-(4); (7)-(8)). Interestingly with loss in years of schooling, one would normally expect a loss in skill if schooling is efficient but this is not the case, as these result in no statistically significant effect of the schooling on the skill of those that have parents with post-secondary schooling. However, being born in quarter 4 to parents with no post-secondary education explains a rise of at least 1.5 years of district-level schooling statistically significant at the 0.1% level (columns (5)-(6); (9)-(10)). Interestingly, for those that have parents with no post-secondary schooling, in lieu of their individual schooling, the substantial (exogenous) variations in their district-level schooling (attributable to their quarters of birth, and interactions with the reform), together with inefficiency in schooling[74], are arguably what explain the positive and statistically significant effects of schooling on their skill. The evidence suggests a standard rise of at least twenty percentage points in reading proficiency for an additional year of schooling for those who have parents with no post-secondary schooling (see columns (3) and (5) of Table 2.16).

In summary, the evidence on parental education shows some consistency using socioeconomic status. Both the interaction effects of the reform and being born in quarter 4 for those that have both high and low socioeconomic status explain a rise in district-level

[74] The evidence from previous analysis suggests a case where, relative to having high educational attainment, lower categories of schooling or educational attainment raise more skill.

schooling of 2.090 and 0.936 years of schooling statistically significant at the 1% and 5% levels, respectively (In columns (11) – (14) of Table 2.21). However, one can only draw a causal inference from an additional year of schooling for those with low socioeconomic status at age 15 (columns (6)-(7) of Table 2.15) where a standard rise of twenty percentage points in reading proficiency is attributable to an additional year of schooling.

Hence, among those impacted by the reform, the effect of an additional year of schooling on skill from which causal inference is drawn is a standard rise of 31.8 percentage points in reading proficiency. The reform (and associated quarter of birth) results in substantial (adverse) variation in the schooling of those who have parents with post-secondary education and no statistically significant effect of the schooling of these from which causal inference can be drawn. However, for those who have parents with no post-secondary schooling (the disadvantaged), the findings suggest the reform (and associated quarter of birth) raised their average schooling across districts. This further suggests that an additional year of schooling is causal to a standard rise of 20-23.5 percentage points in cognitive skill for this. This is also consistent with the findings using socioeconomic status. Here, a rise in the average schooling across districts regardless of socioeconomic status (high or low) is attributable to the reform (and associated quarter of birth), however, an additional year of schooling for those with low socioeconomic status is causal to a standard rise of about 20 percentage points in cognitive skill and no statistically significant effect of an additional year of schooling on the skill from which causal inference can be drawn. I argue that the positive and statistically significant effects of schooling on the skill of the ‘disadvantaged’ over the ‘advantaged’, is (substantially) attributable to the evidence of substantial inefficiency in schooling in urban Kenya.

Table 2-21 Effects of Instruments (IV) on Schooling — First Stage IV

Interaction of QoB and the P1985_	Pool		Indicator of Father's Post-Secondary Education				Indicator of Mother's Post-Secondary Education				Indicator of Socioeconomic Status			
	(1)	(2)	(3)-Hgh	(4)-High	(5)-Low	(6)-Low	(7)-High	(8)-High	(9)-Low	(10)-Low	(11)-Low	(12)-Low	(13)-High	(14)-High
	YoS	Avg_YoS	YoS	Avg_YoS	YoS	Avg_YoS	YoS	Avg_Yo	YoS	Avg_YoS	YoS	Avg_Yo S	YoS	Avg_Yo S
0.p1985_#1.qob_4	-0.040	1.334***	-3.200	-1.834	1.802	2.023***	-4.667	-2.432	1.497	1.505***	3.114	1.920*	-2.357	0.830
	(0.950)	(0.000)	(0.275)	(0.212)	(0.139)	(0.000)	(0.109)	(0.096)	(0.170)	(0.001)	(0.092)	(0.017)	(0.100)	(0.093)
1.p1985_#0.qob_4	1.658***	1.129***	-3.542**	-1.178	0.641	1.427***	-4.454**	-1.681*	0.124	1.060**	1.444	1.554**	1.289	0.600
	(0.000)	(0.000)	(0.005)	(0.060)	(0.493)	(0.000)	(0.004)	(0.028)	(0.875)	(0.001)	(0.218)	(0.002)	(0.176)	(0.076)
1.p1985_#1.qob_4	1.925***	1.425***	-3.440*	-1.137	0.971	1.860***	-4.268**	-1.572	0.285	1.368***	1.837	2.090**	1.478	0.936*
	(0.000)	(0.000)	(0.010)	(0.089)	(0.376)	(0.000)	(0.009)	(0.054)	(0.751)	(0.000)	(0.206)	(0.001)	(0.202)	(0.023)
0.p1985_#1.qob_3	-0.085	0.036	-0.343	-1.490	-1.979	-0.599	-0.667	-1.952*	-1.440	-0.516	1.429	1.394	-0.530	-0.596
	(0.888)	(0.874)	(0.827)	(0.058)	(0.111)	(0.246)	(0.717)	(0.035)	(0.187)	(0.247)	(0.383)	(0.052)	(0.654)	(0.156)
1.p1985_#0.qob_3	-0.586**	-0.140	-0.259	-0.050	-0.541	-0.017	-0.360	-0.089	-0.350	0.014	0.250	-0.081	-0.527	-0.268
	(0.004)	(0.069)	(0.398)	(0.741)	(0.168)	(0.916)	(0.339)	(0.637)	(0.243)	(0.908)	(0.672)	(0.752)	(0.256)	(0.102)
0.p1985_#1.qob_2	-0.042	0.151	-4.771**	-1.969*	-0.534	0.207	-6.67***	-2.913**	-0.065	0.171	0.964	-0.155	-0.594	-0.090
	(0.938)	(0.458)	(0.002)	(0.012)	(0.619)	(0.643)	(0.000)	(0.002)	(0.944)	(0.653)	(0.456)	(0.783)	(0.582)	(0.814)
1.p1985_#0.qob_2	-0.192	-0.151*	0.0244	-0.047	-0.105	-0.093	-0.277	-0.058	0.179	0.025	-0.676	-0.368	0.200	-0.123
	(0.330)	(0.043)	(0.935)	(0.751)	(0.791)	(0.571)	(0.438)	(0.748)	(0.552)	(0.836)	(0.223)	(0.128)	(0.659)	(0.442)
_cons	10.02***	9.524***	17.20***	12.29***	10.06***	8.846***	18.67***	12.89***	10.57***	9.311***	11.29***	9.327***	8.630***	10.00***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	3061	3061	646	646	689	689	375	375	1109	1109	351	351	697	697
R-sq	0.012	0.037	0.034	0.011	0.017	0.095	0.072	0.028	0.009	0.056	0.019	0.081	0.027	0.028
adj. R-sq	0.010	0.035	0.024	0.000	0.007	0.086	0.054	0.010	0.002	0.050	0.001	0.063	0.017	0.018

Note: Table reports full outputs of the first stage of the 2SLS-IV instrumenting individual and average schooling across districts using the interaction of quarter of birth and the reform indicator. Please, see Table 2.15 for the second-stage equation. Columns (1)-(2) refers to the pool, columns (3)- (6) are subsamples based on the father's education; columns (7)-(10) are subsamples based on the mother's education; columns (11) - (14) are subsamples based on socioeconomic status. For each of the measures of background characteristics ((3)-(14)) refer to the subcategories High and Low. The outcome variables are Schooling (individual) as YoS; and Average Schooling (District), as Avg_Yos. The inputs/predictors are interactions of each quarter of birth (QoB) and the reform indicator (p1985_). No variables are controlled for in the model. The p-values in parentheses: * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

2.4 Summary and Concluding Remarks

In summary, the main objectives of this study entail examining the effects of background characteristics on schooling and skill; and an assessment of impacts of education on skills. To improve the internal validity of estimates, I exploit the 1985 curriculum reform in Kenya. The main finding suggests that the reform (and associated instrument, such as Quarter of Birth) result in (exogenous) substantial variation in schooling and skill. The evidence suggests that the effects of the reform on skill is strongly mediated by schooling. In other words, the effects of the reform on skill come through the effect of the reform on schooling. As a response to the first research question—effects of background characteristics on schooling and skill—exploiting the reform in a quasi-experiment (DiD), the evidence suggests, in a similar fashion as the effect of the reform on schooling and skill, the reform-affected estimates of background characteristics suggest, background characteristics impact skill through their effects on schooling.

Contrary to the baseline (OLS) estimates of the effects of background characteristics on schooling and skill, the reform-affected (ATET) estimates of the effects of background characteristics suggest evidence of downward educational mobility for the 'advantaged' and upward educational mobility for the 'disadvantaged'. However, the evidence from the Difference-in-Differences (DiD) shows no evidence of the effects of background characteristics on skill even with the substantial evidence of the effect on schooling that is robust to the use of all measures of background characteristics in this study. I argue that this is due to the limitations of the ATET (Average Treatment Effect on the Treated) and the dependence on the effects of background characteristics on skill (on schooling). Hence dependence on the effects of background characteristics (on skill) on the efficiency of schooling in urban Kenya. To manage these defects or shortcomings, I test for the efficiency of schooling in urban Kenya setting the scene for the response to research question 2—the effects of schooling on skill—proposing a different estimator (2SLS-IV) that gives the LATE to overcome the shortcomings of the ATET.

Findings suggest material inefficiency in schooling in urban Kenya. Further assessing or decomposing the effects of schooling on skill using the Oaxaca-Blinder decomposition suggests substantial inequality and inequity in skills based on background characteristics (parental education and socioeconomic status). However, whilst inequality in skill is attributable to school characteristics (years of schooling), the inequity in skill is not attributable to differences in any observed return effects. Finally, to draw causal inferences from the effects of schooling on skill.

The evidence suggests an additional year of schooling in urban Kenya is causal to a standard rise of 31.7 percentage points in reading proficiency, which strongly suggests schooling substantially raises useful skills in urban Kenya. However, further heterogeneity suggests, that the reform-affected (LATE) estimate of the ‘disadvantaged’ is a standard rise of (20-23) percentage points in reading proficiency for an additional year of schooling with no statistically significant effect of background characteristics on which causal inference can be drawn for the ‘advantaged’. The latter is known to be highly educated relative to the former (see nonparametric evidence). Although robust, consistent, and complementary in several aspects with the effects of background characteristics on schooling using the ATET, these LATE (and ATET) estimates are ‘inherently’ subject to the substantial inefficiency in schooling with evidence of the reforms targeted at the ‘disadvantaged’ at the detriment of the ‘advantaged’[75]. Hence, a consequence of schooling reforms that lack support for useful educational attainment and a highly skilled workforce is crucial for growth and development.

Central to the equity-quality argument raised in the introduction[76] as the basis of the set research questions—the effects of background characteristics on schooling and skills for equity consideration, and the effect of schooling on skills for quality consideration. The substantial evidence of inequality in skill is attributable to differences or gaps in schooling characteristics and not potential discrimination (consistent with the inefficiency in schooling) between the ‘advantaged’ and the ‘disadvantaged’. This may suggest more emphasis on equity (increasing schooling for all) over quality is of a noble aim. However, the substantial inefficiency (poor quality) in schooling where skills from lower levels of schooling is greater than those of higher levels of schooling further suggests efforts or reforms aimed at raising schooling (beyond isced2) for skills will result in unproductivity

[75] This is reflected in the estimates of the ATET and LATE. Specifically, the consistency and complementarity of the LATE which is deemed robust overcoming some defects of ATET. Particularly, for the disadvantaged, the estimates of the LATE is consistent, complementary and robust to the ATET as the positive effects of background characteristics on schooling using the ATET is evidenced on the positive effects of schooling on skill for the disadvantaged using the LATE overcoming the defects of the ATET that suggested a nil effect of background characteristics on skill for the disadvantaged. Similarly, for the advantaged, the nil effect of background characteristics on skill amidst the adverse effect on schooling using the ATET is consistent to the LATE that suggests nil effects of the schooling on the skill of the advantaged. The effects of the latter (advantaged) although consistent using the ATET and the LATE still present issues to resolved, although these are argued to the subject to the inefficiency in schooling where schooling at higher levels (which may characterise in the advantaged) may result in loss of skill or possibly no skill. However, this will require further analysis.

[76] In the introduction, useful balance for equity and quality in schooling is argued — the former is argued to be crucial given the nature of SSA.

(hence, waste of resources) in skill with current quality levels. This suggests the need to raise skills or quality from schooling post-isced2 before further efforts towards equity/equality efforts that raise access to schooling for increased or equitable skills regardless of background characteristics.

To reiterate a few limitations pointed out earlier. Although this study provides an identification strategy used to draw causal inferences on some key estimates. However, a few are left out, whilst this presents a limitation (on the internal validity of few estimates) of this study, I argue that these OLS estimates especially for the effects of the instruments on schooling and skill (see F-tests) are considerably robust, accounting for strata effects, including human capital externality, with clustered robust standard error at the district level (see Equation 2.6). Specifically, to obtain useful estimates of the effects of background characteristics on schooling and skill (research question 1); and in turn, the effects of schooling (continuous schooling) on skill (research question 2).

However, the basis on which the inference on the efficiency of schooling (research question 2—basis of quality in schooling); and the use of the Oaxaca-Blinder decomposition (basis of equality) is deemed not to be causally identified. These two aspects of this analysis are central to the argument on equity and quality as argued in this study. The OLS model estimations (used for both parts of this analysis) are robust. However, possible biases, particularly, the Omitted Variable Bias cannot be completely overlooked. Another limitation of this analysis concerns the design. There is room for future related research to overcome the identification issues in this study. An example would be a strategy that causally identifies the effects of categories of schooling on skills (efficiency of schooling analysis) and; causally identifies the decomposition that explains the skill gap (inequality in skills). There is also room to raise the external validity by considering broader datasets across jurisdictions using longitudinal data that can track individuals over time, particularly, accounting for factors or variables (observable and unobservable) that better (sufficiently) predict the skills of adults in the labour market. This may include data that show linkages between an individual's human capital and an employer's hiring and compensation. Finally, further relationships involving schooling and skill, particularly, how they relate to earnings and the related heterogeneity analysis involving broad skill beyond cognitive skill, education and earnings may further unravel useful insights. In line with this need, the next chapters consider the effects of schooling and skills on earnings (chapter 3); and externalities of schooling and skills on earnings (chapter 4). Finally, chapter 5 considers these key variables—skill, earnings, and education—in a dynamic framework (technology of skill formation).

3 Private Returns to Education and Skill in Kenya

3.1 Introduction

I now examine the pecuniary returns to education and skill (as measures of human capital) of the employed in urban Kenya.

3.1.1 How Productive is Human Capital in the non-OECDs?

Amidst the rise in educational attainment in the non-OECDs[77] (Barro and Lee, 2013), evidence from the Human Capital Index[78] of most non-OECDs still suggests, plans for economic growth and development in related countries are seemingly far-fetched. As of 2018, the HCI of most of the non-OECDs remains below 0.5 (see, World Bank, 2018a). This suggests that if the current levels of investment in education and health persist in the non-OECDs, there is a high risk that a child born today only achieves less than fifty per cent (50%) of their attainable[79] productivity in adulthood (see, Kraay, 2019). The study of Kraay (2019) asserts that there is a significant risk that children born in such regions carry the lasting effects of poor education and malnutrition that adversely impact their cognitive and physical abilities as working adults.

According to the World Bank report (2018) (also see, Corral, et al., 2021 for an update), as of 2018, whilst the HCI for the UK was 0.78, Singapore, Japan, and Korea top the list with 0.88, 0.84, and 0.84 respectively. Colombia recently joined the OECDs and has the lowest HCI among the OECDs (0.59). Across regions of the world, sub-Saharan Africa has the least HCI. Across countries in sub-Saharan Africa, the least on the list is Chad, with an HCI of 0.29, for Ghana, it is 0.44. Besides Mauritius, a small nation, Kenya remains the only relatively large country in sub-Saharan Africa with an HCI slightly above 0.5 mark, with an HCI of 0.52. The HCI is deemed a useful basis for appraising the current provisions of education and health across the countries of the world. The weak human capital in the non-OECDs has been of policy concern not just to the non-OECDs themselves but to other

---

[77] OECD (Organisation for Economic Co-operation and Development) is made up of thirty-eight member countries, founded in 1961. These countries are deemed High-Income. Hence, the non-OECDs make up the world's Low- and Mid-Income Countries which is inclusive of sub-Sahara Africa which comprises Kenya.

[78] The HCI combines indicators of health and education (quality and quantity) into a measure of human capital. Please, see Kraay (2019) for a guide to the HCI methodology.

[79] Attainable productivity is the benchmarked productivity that comes with the level of investments in education and health that gives full education (quality and quantity) and health. Please, see Kraay (2019).

International Development agencies that have continued to advocate and support ‘more schooling’ as an intervention for poverty alleviation, improved health outcomes in most of the non-OECDs (Sayed and Ahmed, 2015; World Bank, 2018b). It is argued that a crucial purpose of the HCI is to incentivise (tenured) policymakers, who due to the long-time horizon involved, from investments in human capital to its vesting, may be less motivated to maintain good levels of investment in human capital—health and education—that will support useful productivity. However, the methodology of the HCI assumes a certain (or static) prevailing level of investment in the quality and quantity of education in projecting future levels of productivity. Over time, the levels of investment in schooling (which substantially influence the quality and quantity of schooling; and health outcomes) may change. This is the case in Kenya, where over time, the changes in the levels of investment in schooling, evident from changes in schooling reforms; and evidence of the change in the quality of schooling, loom large (see Oledibe, 2023a (forthcoming)). Suggesting the current predictions of the HCI may be far from what it appears to be. Besides this (apparent) weakness of the HCI in predicting useful future productivity with the current levels of investment in human capital. The idea of the quality and quantity of schooling in assessing the productivity of (or returns to) human capital in the non-OECDs is plausible and stimulating acknowledging some long-standing arguments that pertain to human capital—schooling and skills—and productivity, in developing contexts.

Several studies (across the literature on the Economics of Education and Labour) have considered individual skill possessed and the years of schooling (or credential) attained as measures of quality and quantity of schooling attained (see Hanushek and Zhang, 2009) that determine useful human capital. As opposed to the mere returns to the ‘quantity’ of education (years of schooling) which is the norm in the literature on returns to schooling, estimates of the pecuniary returns that assess the returns attributable to the quality of schooling attained consider outcomes of schooling (skills) that directly relate to productivity in employment. In this study, I argue that examining the effects or returns to both dimensions of human capital—quality (as skills) and quantity (as years of schooling) of schooling—in estimating the productivity of human capital not only overcomes the weaknesses of HCI that assumes a (certain) constant level of investment but doing this aids useful understanding of returns to human capital and may unravel useful insights in developing contexts contributing to the debate on schooling, skills, productivity and growth in the non-OECDs, ultimately, improve understanding on arguments that suggest schooling give little or no skills deemed to raise productivity (see Pritchett, 2001).

Furthermore, the emergence of the Human Capital theory made estimating returns to schooling a mainstream area of research for economists. Estimates of (private) return to education and skills are used to summarise the economic benefits of investments in schooling (or human capital). In line with the human capital theory, individuals and policymakers are deemed to be fully aware of returns to schooling or human capital and actively decide the optimal levels of schooling with their knowledge of return estimates. Moreover, for the government, understanding estimates of the returns to schooling makes it possible to set (via policymaking) thresholds in schooling (and skill levels) to achieve a desired level of human capital that sustains a useful level of skills, earnings, and employment for growth and development aims. Finally, a useful understanding of the pecuniary returns to years of schooling (across credential categories); and quality of schooling (skills from schooling) attained further unravel insights on the pecuniary effects of any inefficiency[80] and inequality[81] in the schooling of individuals in urban Kenya. This improves understanding of the link between education and skill (see Chapter 2). Hence, the overarching question/objective of this study is:

What is the return to education and skill (as measures of human capital) in urban Kenya?

Beyond providing estimates from which causal inferences are drawn, I examine estimates across genders and employment categories further decomposing differentials in returns to effects of possible inequality and potential discrimination across the gender and employment subsamples.

3.1.1.1 Some Frontiers and Empirical Approaches

In the literature, the idea that schooling at best, yields minimal skills in developing contexts, relative to developed contexts (see Pritchett (2001)) has continued to gain grounds. In part, this is the basis for the use of the qualitative measures of schooling (test scores) in a growing strand of the literature (see Hanushek and Woessmann (2015); and Hanushek et al. (2015)) on returns to education. However, at the other extreme, is the strand of the literature which is the mainstream that follows the early study of Psacharopoulos and Arriagada (1986) that emphasises years of schooling as a plausible measure of human capital amidst several other measures including enrolment used at that time. The years of schooling measure remains

[80] Inefficiency in schooling (as in Chapter 2) is the extent to which schooling give skill.

[81] Inequality in schooling (as in Chapter 2) is simply the effects of how differences in schooling attained impact differences in skills acquired across subsamples or groups of interest.

useful for many reasons, and the most important of these is that it aids comparability across regions of the world, particularly in OECD contexts where very robust estimates based on years of schooling exist. The years of schooling measure is underpinned by models of the human capital framework pioneered by Becker (1964). Several researchers have attempted to adapt the Mincerian Wage Equation[82] (Mincer 1974) using other measures of human capital. However, based on the theoretical justification of the Mincerian Wage Equation, the *years of schooling* measure remains most supported by the models used in the typical analysis of returns to education known to follow the human capital framework. However, the outcomes of return estimates that use years of schooling, give effects of average schooling. One would expect some variability in returns to years of schooling across credential categories, which is partly, the argument of ‘credentialists’. This suggests that, beyond the effects of mere years of schooling, returns to credentials may fully account for signalling effects in returns to schooling. Given the importance of these measures of schooling—years of schooling and credentials—I examine return estimates using both quantitative measures of schooling as aforementioned, in addition to the qualitative measure of schooling—reading proficiency that captures cognitive skills (as ‘skill’) deemed to better relate to productivity as earlier argued (see motivation). In addition to cognitive skills, also examined, are the effects of non-cognitive skills or personality traits known to impact earnings in developed contexts.

(1). For this study, as a baseline of estimates of returns to schooling and skill, I explore a suitable variant of the Mincerian Wage Equation (with a slightly different specification that accounts for district-level schooling and skill) using Ordinary Least Squares. The study by Moretti (2006) in the United States of America suggests accounting for district-level human capital improves estimates of private returns. The findings of Kimenyi et al. (2006) in Kenya also concurred with this. Hence, I account for measures of average schooling across districts. Furthermore, in line with the emphasis on skills in this study, the average skills across districts are accounted for, for improved estimates of the returns to skills. In this chapter, rather than examining the direct effects of district-level (average) schooling and skill on earnings[83], the interests is in assessing the inclusion of these district-level variables (as controls) on the return estimates of the schooling and skills of individuals (respondents). This is deemed to account for the wage effects of educational expansion and skill

[82] Please, see the theoretical justification of the Mincerian Wage Equation in the Appendices (A2.1).

[83] The effects of district-level average schooling and skill on earnings in the specification with individual schooling and skill give the externality of human capital (this is examined in the subsequent Chapter).

proliferation (see the study of Araki, 2020) in the districts. Other individual characteristics controlled for include health, using the Body Mass Index (BMI). Typical of studies on returns to human capital, I account for genders and age.

(2) Relative to years of schooling, skills are deemed better measures of human capital (Bassi et al., 2012; Levy and Schady, 2013), particularly, argued to be true in the non-OECDs, where skill levels are seldom ascertained from educational attainment, due to the understanding of the quality of public education, where it is suggested that schooling gives little or no human capital (Pritchett 2001). As earlier mentioned, 'skill' as used in this study, is in two dimensions—cognitive and non-cognitive skills. Behavioural and Personality Psychologists are known to have extensively discussed skills categorised as cognitive and non-cognitive skills. In the related literature on the returns to skills, economists have relied on the skills categorisation of Psychologists. Cognitive skills entail aptitudes for problem-solving, comprehension and reasoning, mainly, proxied by maths, reading, writing and problem-solving skills; on the other hand, socioemotional or non-cognitive skills entail socioemotional behaviours and personality traits that define, reactions, feelings, and behaviours in people. In this study, in addition to examining and providing estimates of returns to measures of schooling (years of schooling; and credentials), I examine and provide useful estimates of returns to cognitive and non-cognitive skills for the employed in Kenya, providing useful evidence required in complementing and comparing existing estimates of returns in the OECDs and few other non-OECDs. Whilst I examine the wage effects of the measures of schooling and skills firstly on the main analytical sample, I then examine heterogeneity in returns across gender and employment[84] subcategories. I apply the Oaxaca-Blinder Decomposition methodology (see details in the Methods Subsections of Chapter 2) across gender and employment categories to examine wage effects due to differences in characteristics (or endowments) that give insights into inequality; and differences in returns (or coefficients) that provide insights into potential discrimination, across each of the subgroups of interest.

(3). Lastly, the endogeneity biases, particularly, due to observed factors such as ability that mars return estimates are expected from baseline (OLS) models warrant raising the validity of the estimates in the intensive margins by pursuing an identification strategy. A useful identification strategy supports estimates of returns to schooling and skills (cognitive) from

[84] Specific categories of employment considered are the informal/formal, the public/private service wage-employed and those in entrepreneurship/lone employment.

which causal inferences are drawn. I deploy an identification strategy inspired by the study of Harmon and Walker (1995) in the United Kingdom. Harmon and Walker (1995) propose a useful approach to obtaining estimates from which causal inference is drawn by exploiting exogenous variations in schooling attributable to a schooling reform—that raises the minimum school leaving age in the UK from 14-15. In this study, by exploiting exogenous variations in schooling and skill attributable to curriculum (structural) reforms in Kenya (see Chapter 2 for the implementation of the reform and the specification of the reform indicator) using the 2SLS-IV (the Two-Stage Least Squares Instrumental Variables) approach. The evidence from the first-stage equation suggests that the 1985 curriculum reform (and associated Quarters of Birth) resulted in exogenous variations in schooling and skills. Hence, these provide useful instruments to obtain estimates of the 'causal' wage effects of measures of human capital of interest. This makes it possible to go beyond descriptive evidence (OLS) to obtain return estimates that strongly impact policymaking whilst appraising the reform as the causal estimates obtained are deemed Local Average Treatment Effect (LATE) that relate to those impacted by the reform. While the 2SLS-IV approach to drawing causal inference relates to the effects of years of schooling and cognitive skill, I argue strongly that the OLS return estimates of non-cognitive skills are sufficiently robust and causal inference can be drawn from estimates of the returns to non-cognitive skills.

Hence, in responding to the overarching question, I respond to the following sub-questions as thus:

1. What is the return to measures of human capital (schooling and skills) for the employed in urban Kenya? (1) and (2).

2. What is the reform-affected return (or the causal effects) of education and skill on wages? (3)

3.1.2 Antecedents and Contributions

Antecedents

Within the human capital framework (Schultz, 1961; Becker, 1962; Mincer, 1958 and 1974), it is known that wage differences for the employed arise through differences in their human capital accumulated through schooling; and on-the-job training or experience. While this may suggest those with high education and (or work experience) have high human capital and hence, high skills for their (possible) higher earnings (wage returns). This is not always the case, particularly in the non-OECDs. Besides arguments that high educational attainment does not necessarily translate to high skills (Hanushek and Woessmann, 2015; Pritchett 2001), particularly, in the non-OECDs, the empirical evidence suggests, those with high returns are not necessarily productive or with high skills (see Serneels, 2008). This is no doubt at odds with the human capital theory that suggests, returns are to come through productivity which requires useful skill from schooling. Serneels (2008) who used the Methodology of Medoff and Abraham (1980) in the U.S. for Ghana, finds that wage seniority, although dependent on education, is independent of performance or productivity. Ultimately, Serneels (2008) provide evidence that suggests returns to schooling bear no relationship to productivity. However, the returns to cognitive ability relate to productivity. Suggesting measures of cognitive skills better capture human capital in the non-OECDs as suggested by Hanushek and Woessmann (2015). Whereas the evidence suggests schooling strongly explains seniority, hence earnings. However, in this study, rather than focusing on cognitive skills as a sole measure of human capital, I examine comprehensively, both the effects of schooling (years of schooling and credentials) and skills (cognitive and non-cognitive) in Kenya. I argue that, in addition to measures of skills, examining schooling through which skill (besides experience, age or other measures that impact skill) is raised is useful. This is useful because, the wage effects of the former may relate to the wage effects of the latter in a way that can unravel useful policy insights (see Oledibe, 2023a (forthcoming). In the previous chapter, the evidence in Kenya suggests that schooling explains skills amidst substantial evidence of inefficiency in schooling in urban Kenya.

Furthermore, in this study, the emphasis on skills both as a measure of the quality of schooling and a measure of human capital in the non-OECDs makes discussing the literature on returns to the schooling of immigrants (or by race or quality of schooling received) (Card and Krueger 1992a; Card and Krueger 1992b; and Bratsberg and Terrell 2002) interesting. Specifically, the work of Schoellman (2012) sets the scene for a useful review. Schoellman (2012) shows that the returns to the schooling of foreign-educated immigrants in the US do not only capture the effects of the relative productivity (output per worker) of the employed in the U.S. but also the relative quality of education of the respective countries of the immigrants. Findings suggest, that in accounting for productivity, cross-country differences in years of schooling are almost as important as cross-country differences in education quality. Although the findings, models used, and variable specifications differ substantially from the approach of this study, several interesting methodological contributions relate to this study specific to urban Kenya and the non-OECDs. First, although the literature focuses on the use of Development Accounting which underpins the HCI (earlier discussed) methodology and differs from the reduced-form approach of this study. The evidence from Schoellman (2012) suggests that estimates of returns to schooling differ significantly between immigrants of OECDs and non-OECDs in the United States of America, using the U.S. labour market as the base (or reference of this study). This suggests differences in the quality of schooling in the home countries of the migrants not only accentuate the differences in the levels of skills from their schooling but also show some consistency with the understanding of the quality of schooling across OECDs and non-OECDs. Again, this finding is consistent with Pritchett (2001) who argued that the lack of productivity in the region may be explained by the lack of skills from schooling which further explains the fast-decreasing returns to education, in the non-OECDs.

The baseline findings from Schoellman (2012) suggest that the differences in education quality and years of schooling alike account for the productivity of workers and are consistent with earlier findings and arguments (Pritchett 2001; Hanushek et al., 2015) that suggest, poor-quality schooling yields low skills which explains low productivity and high-quality schooling yields high skills and explains high productivity. In addition to differences in wages (of migrants of OECD and non-OECD origins) attributable to the quality of schooling received. although educational attainment is on the rise, in the non-OECDs, the average schooling in the OECDs surpasses the average schooling in the non-OECDs (Barro and Lee, 2013). This further supports the claims on differences in schooling accounting for productivity across the OECDs and the non-OECDs even without accounting for the quality of schooling. Finally, Schoellman (2012) finds that the quality of schooling best explains productivity as the quality-adjusted years of schooling best account for output per worker. Interestingly, like the findings of Hall and Jones (1999), the findings of Schoellman (2012) suggest replacing education quality (as defined) and years of schooling of immigrants with the education quality and years of schooling of the U.S. raises output per worker in the range (from 3% to 20%). This finding suggests that the quality of schooling substantially explains skills, which in turn, raise the return/productivity of workers. This gives further credence to the emphasis on skills in this study, as a measure of the quality of schooling attained that directly relates to the productivity of the employed. Furthermore, the study of Schoellman (2012) further inspires an analysis of the differences in returns to the schooling and skills of categories of the employed to further assess the extent to which potential discrimination and inequality in schooling and skills explain the differences in returns to schooling and skills of categories of genders and the employed of interest in this study.

Furthermore, the work of Boissiere et al. (1985) is one of the relatively few studies on returns to skills in non-OECDs that provides evidence from Kenya. With the use of a recursive wage specification with structures. Although the methods used suggest findings may be credible. However, the complex (and unrealistic) structures have conditions that deviate from the current understanding of schooling, human capital, and earnings for Kenya. Firstly, the methods in Boissiere et al. (1985) suggest that years of schooling alone must impact human capital which then determines earnings—this condition unrealistically suggests, that the completion of secondary schooling alone fully develops cognitive skills that raise human capital, and, in turn, raises earnings. Next, a further condition in the structures suggests, unrealistically, that completion of secondary schooling, which is a proxy of cognitive skills, is completely exogenous. Suggesting ability has no indirect (through completion of secondary schooling) or direct effects on earnings. What is clear in this (Boissiere et al. (1985)) interesting study is that whilst the conditions in place attempt to overcome issues of data availability, especially for studies in the non-OECDs that only recently have data that capture cognitive abilities or skills in the non-OECDs which makes substantial contributions. In this study, rather than complex structures such as Boissiere et al. (1985), I examine cognitive skills and human capital, for the employed, in urban Kenya. I follow a reduced form approach as the study of Ingram and Neumann (2006) in the United States of America. These studies are concordant with the key understanding of human capital theory (Schultz, 1961; Becker, 1962; Mincer, 1958/1974). In addition to the objective to examine skills, going beyond mere years of schooling, this study examines human capital (comprehensively) using years of schooling and reading proficiency, acknowledging the precedence of schooling for skills formation.

In addition to the returns to education, this study focuses on returns to cognitive skills and personality traits for the wage-and self-employed in Kenya. This study aims to provide similar evidence like those of Hartog et al., (2010) and other similar studies for OECDs. Studies of this sort are still rare in non-OECD contexts. The works of Hartog et al. (2010) compared the pecuniary returns to cognitive and 'social abilities' for the wage-employed and entrepreneurs using the Difference-in-differences (DiD) approach, and they found that, relative to social abilities, cognitive skills have a stronger impact on earning for entrepreneurs compared to the wage-employed. Understanding crucial skills for the distinct categories of the employed will strengthen strategies for full employment or workforce planning in sub-Sahara Africa. The place of skills for growth cannot be over-emphasised. Evidence from the Netherlands reveals that cognitive skills in the forms of mathematical and technical abilities in early childhood better explain earnings for entrepreneurs (the self-employed), while clerical and language abilities better explain earnings for the wage employed (Hartog et al., 2010). For the United States of America, evidence from the works of Levine and Rubinstein (2013) shows that adolescents with high self-esteem, high-order cognitive skills and rule-breaking tendencies are more likely to enjoy successful long-term careers as entrepreneurs in adulthood. What skills best explain earnings in wage- and self-employment in sub-Sahara Africa? Put differently, what skills are most rewarded in the Kenyan labour markets? This study aims to provide similar evidence as cases mentioned in the Netherlands and the United States of America. Specifically, this study focuses on the public service employed within wage-employed, and the entrepreneurs within the self-employed, in Kenya.

Contributions

This study contributes to the literature by improving on a few fronts. Firstly, this study improves understanding by raising the internal validity of return estimates, implementing a useful identification strategy, that helps to draw causal inferences in estimates using the 1985 curriculum reform that shows evidence of an exogenous variation in schooling. I argue that consideration for skill as a measure of the quality of schooling attained is particularly, a useful measure of human capital in the non-OECDs deemed subject to issues of quality or efficiency in schooling (Hanushek, 2008) as is the case of urban Kenya from the findings in Chapter 2. Besides evidence from almost four decades ago by Boissiere et al. (1985), no studies have done this for Kenya in recent times. Next, a mediation analysis shows a useful mechanism through which skill impacts earnings. Findings suggest the return estimates of measures of non-cognitive skills are sufficiently robust, hence causal inferences are drawn from the estimates. Further findings show that, although the baseline (OLS) return estimate for cognitive skill is more statistically significant than the returns to schooling. Further evidence suggests that the impact of skill on earnings is through the effect of skill on education, as education mediates (or substantially moderates) the wage effects of skill in urban Kenya. This not only shows that skill is nurtured by schooling, but this indicates that the reward for skill comes through education. This shows some support for human capital and signalling theories. Suggesting, the labour market rewards the skills of the highly educated. Hence, having a high skill without high education may attract a lower reward. This is consistent with the examination of the skill from schooling, where, on average, those with ISCED2 (post-primary) credentials show evidence of the highest skill (evidence of inefficiency in schooling in urban Kenya), but respondents with ISCED56 (tertiary) credentials have the highest returns for their schooling relative to those with ISCED2 credential category. It is somewhat inconsistent with the human capital theory, showing more consistency with the signalling theory. As earlier highlighted, this study also contributed to existing literature by raising the internal validity of return estimates improving knowledge of estimates of returns by providing reform-affected (using the 2SLS-IV approach) return estimates of the effects of cognitive skill and schooling. This provides a basis to draw causal inferences from return estimates. Turning to the findings from the 2SLS-IV approach—which gives estimates of schooling and cognitive skill from which causal inferences are drawn—findings suggest, at best, relative to the female gender, only males have useful returns to their schooling and their skills in urban Kenya. The Oaxaca-Blinder Decomposition suggests substantial evidence of inequality and potential discrimination in returns to schooling for females relative to their male counterparts in urban Kenya.

Another contribution of this study is a consideration for non-cognitive or socioemotional skills. This study presents estimates of pecuniary returns to certain personality traits (the Big 5) for Kenya. In recent studies (that use longitudinal data) involving eleven OECDs, except for the United Kingdom and Canada, cognitive skills better explain earnings in Switzerland, Denmark, Finland, Iceland, Sweden, and Norway. However, evidence from the growing literature suggests that personality traits are (at least) as important as cognitive skills in explaining earnings in some OECDs, especially in Germany, the Netherlands, and the United States of America.

The remainder of this chapter is organised as follows: In subsection 2.2, I discuss the Data and Methods in subsection 3.2 where the variant of the Mincerian Wage Equation; and the Instrumental Variables approach deployed to draw causal inferences from estimates are discussed. In subsection 3.3, I present and discuss the results and subsection 3.4 includes the summary of the main findings and the concluding remarks.

3.2 Data and Methods

In this study, the World Bank's STEP Household Survey, for Kenya, is used. Methods deployed include the Empirical Framework which makes the Estimation and Identification strategies. I now turn to discuss the Data used for the testable predictions of this study.

3.2.1 Data and Descriptive Evidence

As highlighted in earlier subsections, this study exploits the STEP Household Survey for Kenya. Please, see the Data Subsection of Chapter 2 and the related Appendices. Unlike Chapter 2 where the entire survey sample of 3894 respondents was used. This analysis pertains to the employed (those with earnings) which restricts the survey sample. Hence, the analytical sample for the employed excluding the unpaid employed in family businesses, is about 2355. However, missing observations and data cleaning resulted in the effective use of only 2008 observations, at best. To examine heterogeneity in treatment across the employed, the analytical sample is further restricted, for each of the subsets of the employed. Please, see Figure 3.1 below for the diagrammatic representation of the analytical samples and some subsamples of interest, in this analysis.

For ease of exposition, this Data subsection only presents variable specifications and descriptive evidence used in this analysis which are not described or specified elsewhere. Hence, some variables used in this chapter may have been described (or specified) elsewhere, if any, they are cross-referenced as due.

Sub-samples of the Main Analytical Sample (sub-categories of the employed)

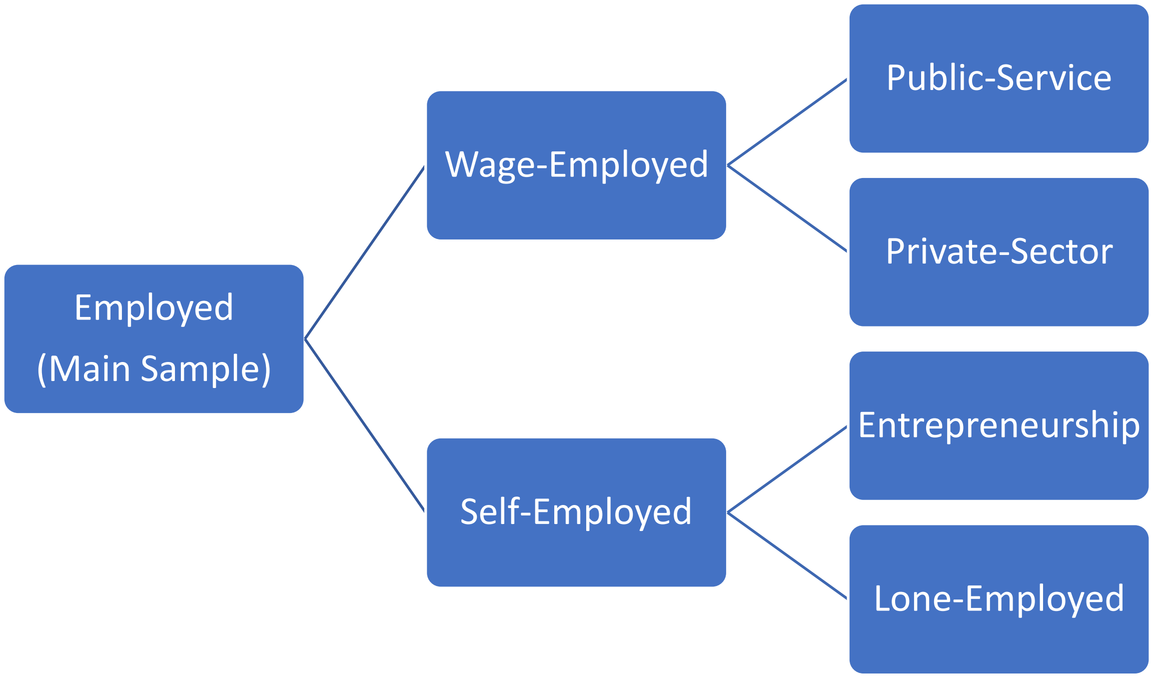


Figure 3-1 Diagrammatic Representation, Subsamples of the Employed

The diagram indicates the self-employed consists of the lone-employed and the entrepreneurs (distinguished by having at least one employee). The wage-employed consists of the private-sector wage-employed and the public-sector wage-employed. Other subsamples used in the heterogeneity analysis include genders (male/female); and informality (the informal/formally employed).

Table 3-1 Variable Description and Some Descriptive Evidence (Summary Statistics)

Variable	Brief Description	Obs	Mean	Std. dev.
earn_h_usd	continuous, hourly earnings in USD.	1,901	4.028	11.062
extrav_av	average of a set of traits that elicit extraversion	2,001	2.877	0.583
consci_avg	average of a set of traits that elicit conscientious	2,001	3.248	0.509
stab_av	average of a set of traits that elicit stability	2,001	2.707	0.499
open_av	average of a set of traits that elicit openness	2,001	2.707	0.499
agree_av	average of a set of traits that elicit agreeableness	2,000	2.868	0.562
BMI	measure of health condition: body mass index	1,930	24.532	4.202
Informal	indicator informality/formality in employment	2,008	0.733	0.443
emp_status	indicator of employment status	2,008	1.413	0.493
pub_emp	indicator of public/private wage employment	1,170	0.125	0.331
bus_size	indicator of lone/entrepreneurship self-employment	826	0.827	0.379
p1985_	reform dummy, instrument for schooling/skill	1,941	0.864	0.343
qob_1_	quarter of birth 1 (reference)	1,941	0.450	0.498
qob_2_	quarter of birth 2, an instrument for schooling/skill	1,941	0.292	0.455
qob_3_	quarter of birth 3, an instrument for schooling/skill	1,941	0.259	0.438
qob_4_	quarter of birth 4, an instrument for schooling/skill	1,941	0.199	0.399

Notes– Source: Author’s elaboration of the STEP Household Survey for urban Kenya. Please, see Data Sections of Chapter 2 and related Appendices for details. This table presents some summary statistics for the sample of the employed, as used in the analytical models (after cleaning).

The survey instruments of the STEP elicited data on respondents’ income, educational attainment, age, gender, labour market and employment statuses. The main income data for the respondents, hourly earnings in USD (earnings_h_usd) is the outcome variable of interest in this study. Specifically, the income data includes hourly earnings useful to the wage-employed and the self-employed. The STEP also provided data on net profits useful for the return estimates of the self-employed—this consists of those in entrepreneurship (conditioned on having at least an employee) and the lone-employed (no employees). These measures of earnings have been carefully constructed in the STEP data. The log hourly earnings (in USD) is the main outcome variable for the wage- and self-employed in this study. I acknowledge susceptibility to biases in estimates. Besides the susceptibility of income under-reporting bias, particularly for the self-employed and informally employed, the use of log of hourly earnings in USD as the outcome variable for models that estimate returns to schooling for both the self- and the wage-employed means further defects of the log-hourly earnings variable may be attributable to the treatment of taxes. Besides the public service wage employed that may have useful tax records with large proportions of the private-sector wage employed and almost all the self-employed being informal accounting for taxes may present substantial income misreporting issues. These issues inhibit comparisons of return estimates between the wage-employed and self-employed. However, rather than marking down or -up (adjusting the stated income by a certain percentage) earnings as some researchers have done, I compare return estimates without adjusting earnings for taxes for the self-employed, however, this is acknowledged in the interpretation of returns between both categories of the employed. This presents a limitation to this study warranting careful interpretations of findings.

3.2.1.1 Specifying the Earnings Variable (In_earnings_h_usd)

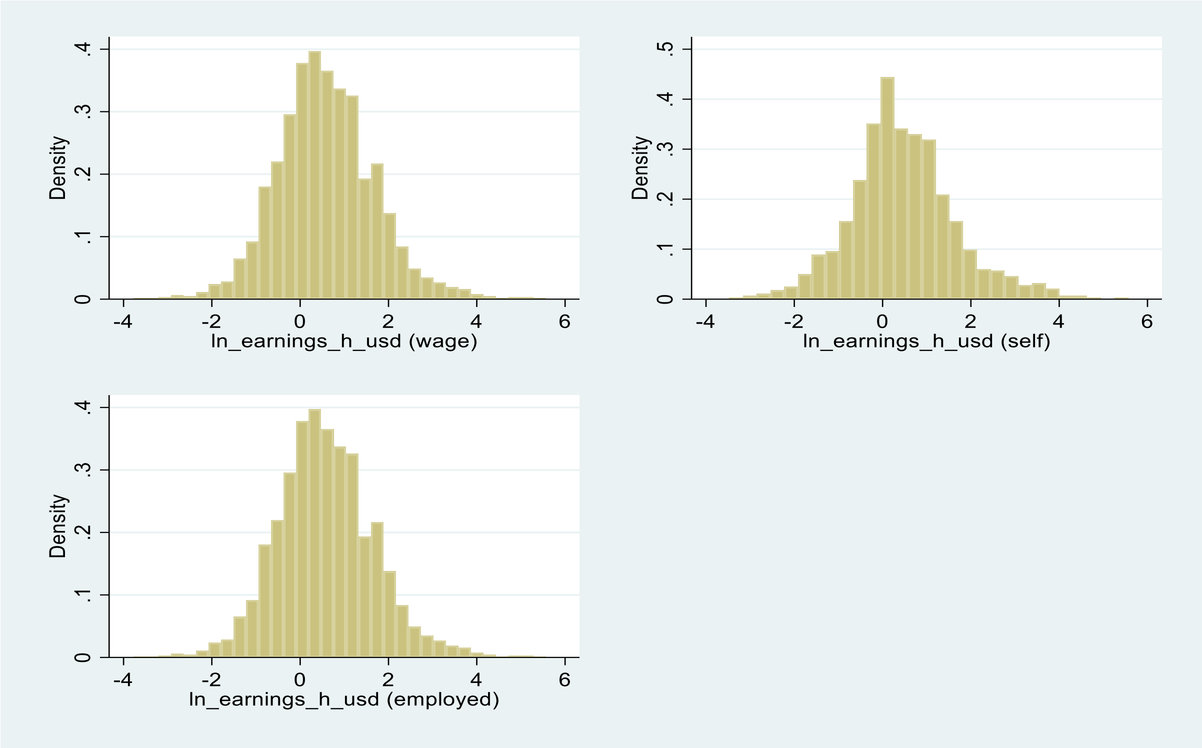


Figure 3-2: Histogram, Log Hourly Earnings (USD), by Key Categories of the Employed.

Source: Author’s elaboration of the STEP Household Survey for urban Kenya.

Taking the logarithm of hourly earnings in USD, *earnings_h_usd* results in the variable, *In_earnings_h_usd*. This normalises the earnings of analytical samples that now have properties that make the best of estimators used in this analysis. Taking the logarithm of (hourly) earnings is about the norm in applied research.

Table 3-2 Descriptive Evidence, Hourly Earnings (USD)

earnings_h_usd (wage-employed)					**earnings_h_usd (self-employed)**			
	Percentiles	Smallest			Percentiles	Smallest		
1%	0.228	0.023			0.145	0.031		
5%	0.407	0.071			0.329	0.043		
10%	0.499	0.071	Obs	1,170	0.483	0.054	Obs	826
25%	0.977	0.109	Sum of wgt.	1,170	0.869	0.071	Sum of wgt.	826
50%	1.827		Mean	3.6321	1.629		Mean	4.069
		Largest	Std. dev.	8.915		Largest	Std. dev.	12.344
75%	3.744	74.877			3.594	80.053		
90%	7.188	143.764	Variance	79.489	7.020	97.714	Variance	152.362
95%	10.483	159.737	Skewness	13.349	11.232	131.336	Skewness	12.647
99%	27.811	175.961	Kurtosis	227.052	38.000	260.571	Kurtosis	227.183

earnings_h_usd (employed)				
Percentiles		Smallest		
1%	0.145	0.023		
5%	0.329	0.031		
10%	0.483	0.043	Obs	1,901
25%	0.869	0.054	Sum of wgt.	1,901
50%	1.629		Mean	3.808
		Largest	Std. dev.	10.437
75%	3.594	143.764		
90%	7.020	159.737	Variance	108.923
95%	11.232	175.961	Skewness	13.425
99%	38.000	260.571	Kurtosis	251.990

Source: Author's elaboration of the STEP Household Survey for urban Kenya.

Using Table 3.2 to compare the hourly wage (in USD) distributions across the wage and self-employed in Urban Kenya show that, the mean and median hourly wages for the employed (main analytical sample), the wage- and self-employed are 3.8075, 3.6321, 4.0685; and 1.6286, 1.8272, 1.4683, respectively. This indicates that while half of the employed in urban Kenya earn above 1.6286 dollars an hour, this is slightly different for the wage and self-employed, where half of the wage-employed earn over 1.8272 (1.4683 for the self-employed) in urban Kenya. Comparing mean hourly earnings shows that the mean hourly earnings of the self-employed are $0.44 (4.0685-3.6321) greater than those of the wage employed. The mean earnings for the employed in urban Kenya is $3.8075. The significantly higher mean earnings over the median earnings for the analytical sample and across the wage- and self-employed subcategories indicate that most employed earn less than their mean earnings discussed, which is characteristic of the skewed distribution in earnings seen in the analytical sample (13.42489) and across the associated subcategories (13.34882 and 12.64694) from Table 3.2. This suggests a distribution with a few extremely high earners and substantial proportions of the employed (particularly the self-employed) earning less than the mean earning across the distribution.

To further consider variations in earnings for the employed and associated subcategories in urban Kenya, there is a need to consider some measures of dispersion, with standard deviations and ranges of 10.4366, 8.915635, 12.3435; and 260.5484, 175.9376, 260.54038 across the employed, the wage- and self-employed, respectively. These indicate that variation in earnings is greatest for the self-employed with a standard deviation and range of 12.3435 and 260.54038, respectively. Variability in earnings for the self-employed gives further insight into possible differences in returns to entrepreneurship and those in lone employment within the self-employed in urban Kenya. The interquartile range is a useful measure in this regard. The interquartile range (75th percentile – 25th percentile) defines the boundary of earnings of the central half of the sample. This gives a better understanding across quarters of the analytical sample and associated subcategories. The 75th, and 25th percentiles and interquartile ranges for the employed, the wage- and self-employed are 3.5941, 3.7438, 3.1234; 0.8686, 0.9771, 0.7505; and 2.7255, 2.7667, 2.3729. From the interquartile range, it is evident that the self-employed show the least variability in earnings between the 75th and the 25th percentile, contrary to the evidence revealed by the range across the analytical sample and subcategories. Comparing the 25th, 50th (median), and 75th percentiles, the evidence suggests the self-employed still have the lowest earnings for all quarters and this continues to the 90th percentile where the self-employed still recorded the least earnings compared to the wage employed. The 95th percentile shows a reversal in earnings as the top 5% of the self-employed earn over $11.2/hr (compared to $10.5/hr for the wage-employed). This further supports the previous argument between substantial earnings/returns for the few entrepreneurs (with substantially high earnings) relative to the numerous low-earning lone employed (with significantly low earnings) most of which may be involved in petty or itinerant trade in urban Kenya.

Evidence from Table 3.6 reveals that the males earn about $0.64/hr more than females on average the variability (SD) in earnings among the males is greater, relative to the females. This evidence suggests that the males earn better than the females in urban Kenya. However, relative to females, on average, the males have five months of work experience; and a year of schooling more.

Evidence from Table 3.7 shows the average earnings across credential categories. The mean earnings for the employed with ISCED34A credential (secondary and some post-secondary) is $2.886038/hr, this happens to be a point that deviates from expectations from the trend between educational attainment and the earnings of the workforce. Members of the workforce with ISCED34A with higher skills have lower hourly earnings relative to workers with lower secondary education attainment (ISCED2). Although the variability (SD) in earnings within the ISCED34A category is relatively low, the ISCED34A is the credential category of most of the employed in Kenya, this may suggest a high supply (without less demand) of the skill type/level of this category of workers in urban Kenya. Interestingly, the higher age and work experience of those with ISCED2 (with an average of under 0.5 years of age; and seven months of work experience, over those with ISCED34A credential category) may explain the deviation of workforce with ISCED34A from the credential-earnings trend in urban Kenya.

Table 3-3 Hourly Earnings (USD), by informal/formal and Age Categories.

	Freq	**%**	**Mean**	**Mean**	**Median**	**SD**	**Min**	**Max**
		Freq	*(Yos)*	*earnings_h_usd*				
Formal	603	24.93	13.419	6.135	3.594	12.366	0.071	159.737
Informal	1,816	75.07	9.553	2.989	1.303	9.539	0.023	260.571
Total	2,419	100	10.514	3.809	1.629	10.439	0.023	260.571
age_group								
15-19	70	2.89	8.171	1.172	0.686	1.540	0.075	11.1674
20-24	509	21.02	10.697	3.557	1.498	11.526	0.107	175.961
25-34	1,078	44.51	11.213	3.685	1.815	7.533	0.054	159.737
35-44	474	19.57	10.034	3.552	1.666	8.250	0.023	131.336
45-64	291	12.01	8.883	5.781	1.721	18.995	0.043	260.571
Total	2,422	100	10.505	3.808	1.629	10.437	0.023	260.571

Source: Author's elaboration of the STEP Household Survey for urban Kenya.

Table 3.3 reveals that the formal workforce earns $3.15/hr more than the informal workforce on average, the median earnings for formal are also significantly greater, relative to those of the informal workforce, although variability (SD) in earnings among the formal is greater relative to that of the informal. The high mean hourly earnings may be explained by the substantial difference (*4 years_educ_act*) in the years of schooling between the formal and the informal workforce. Across the age group, the 45-64 age group also have lower years of education (relative to the 35-44 age group) but significantly higher earnings. This may be explained by a possible (significantly) high years of work experience of the 45-64 age group, relative to the 35-44 age group.

3.2.1.2 Non-Cognitive Skills: Variable Specification and Descriptive Evidence

As earlier mentioned, the measure of non-cognitive or socioemotional skills considered in this study is the Big Five Inventory. This consists of Extraversion, as extraversion_av; Conscientiousness, as conscientiousness_avg; Openness to experience, as openness_av; Agreeableness, as agreeableness_av; Neuroticism (or the opposite, Emotional Stability as stability_av). Table 3.1 presents summary statistics of these traits/skills. A battery of instruments elicits each of these traits. Hence, the traits are composites (averages) of related survey instruments that make up the traits. See Table 2.5 and the related data appendices for the Big-5 personality traits.

These socioemotional skills are known to impact earnings and are beginning to gain more attention, and increasingly ascribed crucial skills for success in the labour market, especially in the OECDs. A unique (and valuable) feature of the STEP surveys is the numerous dimensions of skill it presents. These include the socioemotional skills of respondents. The STEP Household Survey covers these traits relating to (multiple domains of) social, personality, emotional, attitudinal, and behavioural domains. Among other soft skills, these traits reveal the capacity to be creative, manage emotions, work, and relate well with others in workplaces. It has been argued that schooling impacts these traits which are also known to (substantially) impact earnings in employment. As earlier highlighted in the literature review subsection, these skills are known to explain earnings at least to the extent to which cognitive skills do in some countries like the United Kingdom. I examine the impacts of schooling on these traits (in a different study). In this chapter, I examine the wage effects of these traits in urban Kenya. Using Mincer-like earnings functions linked with educational-attainment and educational-production functions; and forming a recursive framework that estimates pecuniary returns to schooling, controlling for employment experience, the work of Boissiere et al. (1985)[85] estimated returns to cognitive skills, native ability, and secondary schooling in Nairobi Kenya. Although Boissiere et al. (1985) have not used the Big 5, they have used other suitable measures of socioemotional skill. Consistent with the findings of Boissiere et al. (1985), is the study of Otchia and Yamada (2019) who used the STEP Household data to explain the effects of heterogeneity of skills in Kenya. Otchia and Yamada (2019) find that cognitive skills have higher returns relative to non-cognitive skills however evidence from their works also suggests that non-cognitive skills explain the earnings of the highly educated/skilled. For the full sample, in Kenya, they found that Conscientiousness and Openness are positively associated with earnings. Agreeableness is negatively associated with earnings (however, positively, associated with the earnings for highly skilled). Otchia and Yamada (2019) also found that controlling schooling reduces the effects of Openness, but Conscientiousness and Agreeableness stay the same. Suggesting schooling may strongly explain the former (Openness) relative to the latter (Conscientiousness and

[85] As earlier highlighted in the Introduction subsection (see antecedents), Boissiere et al. (1985) used completion of secondary education as the schooling variable, and they estimated returns to cognitive skills proxied by literacy and numeracy tests; and they also estimated returns to ability proxied by test scores – Raven's Progressive Matrices – the test entails matching of pictorial patterns which does not require the use of cognitive skills. Findings from the works of Boissiere et al. (1985) reveals that return to cognitive skills are large, relative; and years-of-schooling is moderate in Kenya.

Agreeableness). In this study, I present new estimates of the returns to these traits and compare or critique findings to the related (existing) studies for Kenya, sub-Sahara Africa including the rest of the non-OECDs and the OECDs at large.

Figure 3.3 and Table 3.4 show some descriptive evidence of each trait. In Fig 3.3, the y–axis shows densities; and the x-axis shows a scale (1-4), with 1 for minimal or no evidence of trait and 4, for the highest evidence of trait. Please see Appendix A2 for details of each of the traits. The emotional stability seems quite noticeable among the employed particularly, within most of the other subsets (of interest) of the employed. The skill of stability has the highest peak. This suggests that relative to other traits, a good proportion of the employed shows mild (2-3) evidence of the trait of emotional stability. Table 3.4 shows that regardless of educational attainment, most respondents exhibit moderate levels of most traits. The trait of openness (for respondents with post-secondary/advanced and non-tertiary qualification, the ISCED 4B) has high peaks between (3 - 4) scales, suggesting that a useful proportion of respondents with the ISCED4B credential (relative to other credential categories) shows substantial evidence of the trait of openness. This is expected, as this credential category should make up part of those in technical occupations, having specialised qualifications for their jobs but a closer look at Fig 3.3 shows that those with a university degree, the ISCED56 credential category, have higher skills of openness, relative to those with ISCED4B. Further evidence suggests that those with the ISCED4B credentials have a higher trait of conscientiousness and emotional stability relative to other credential categories.

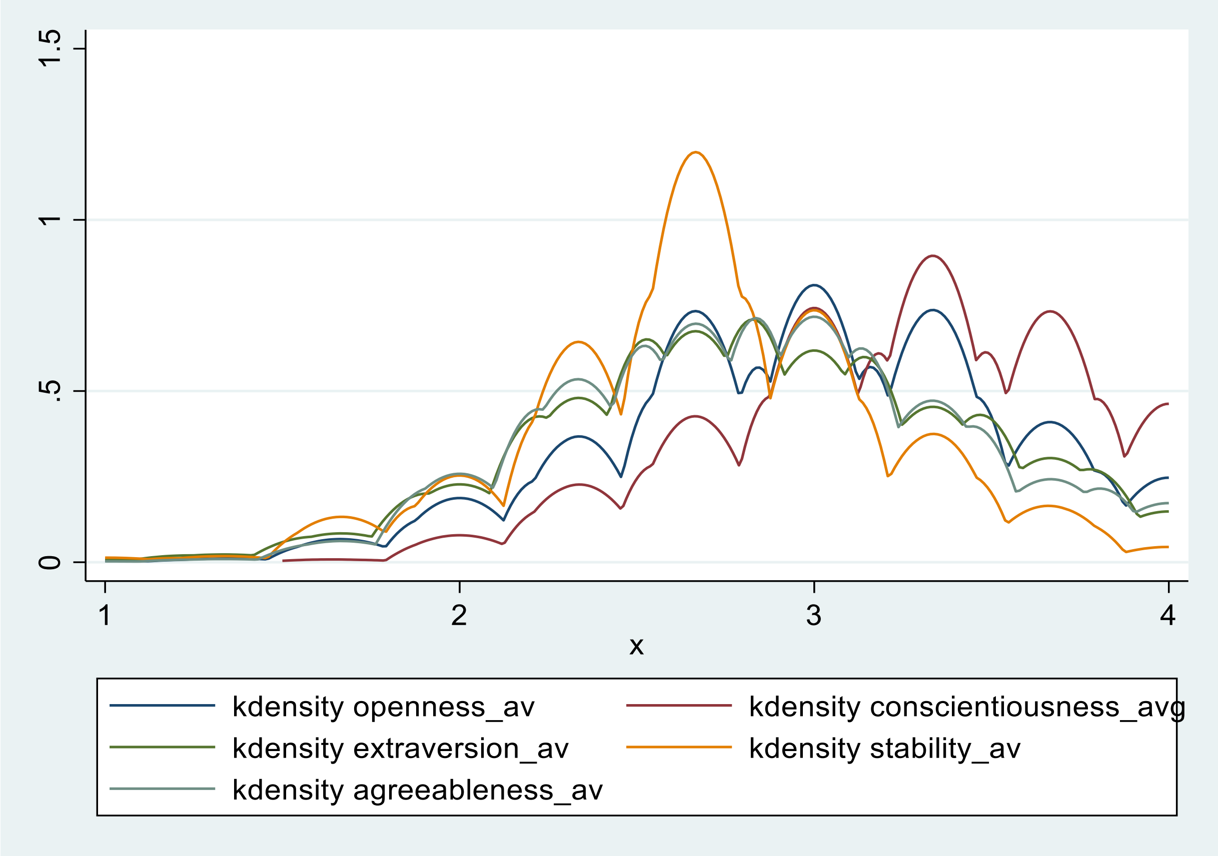


Figure 3-3: Kernel Density Plot, Personality Traits (Big 5).

Source: Author's elaboration of the STEP Household Survey for urban Kenya.

Table 3-4 Personality Traits (Big 5), by ISCED, Employment, Age and Genders

	Agree_av	Stab_av	Ext_av	Open_av	Cons_av
ISCED01	2.840	2.695	2.750	2.867	3.204
ISCED1	2.819	2.693	2.769	2.844	3.170
ISCED2	2.869	2.708	2.759	2.893	3.259
ISCED34A	2.842	2.723	2.922	3.017	3.257
ISCED4B	2.903	2.727	2.936	3.196	3.314
ISCED56	2.909	2.723	2.952	3.227	3.293
Formal	2.900	2.748	2.924	3.124	3.293
Informal	2.840	2.698	2.832	2.947	3.225
Wage	2.841	2.730	2.857	2.994	3.244
Self	2.885	2.685	2.864	2.988	3.244
Unpaid	2.710	2.710	2.672	2.989	3.156
Male	2.854	2.743	2.864	3.033	3.263
Female	2.857	2.670	2.845	2.939	3.215
1 (15-19)	2.633	2.676	2.821	2.889	3.106
2 (20-24)	2.805	2.691	2.826	3.027	3.192
3 (25-34)	2.850	2.710	2.906	3.026	3.230
4 (35- 44)	2.896	2.704	2.812	2.923	3.278
5 (45-64)	2.950	2.770	2.796	2.939	3.349

Source: Author's elaboration of the STEP Household Survey for urban Kenya.

While those with ISCED56 show evidence of high Agreeableness (relative to other categories), this may suggest that being agreeable is well rewarded for the highly educated in Kenya. Those with ISCED34A show evidence of higher extraversion. The self-employed show higher (average) agreeableness and extraversion, this is consistent as owning a business should have some association with useful communication to thrive. However, the wage-employed show evidence of higher (average) traits of openness, stability, and conscientiousness. While the females show higher (average) trait of agreeableness, the males surpass the females on the rest of the personality traits. The evidence suggests that agreeableness, conscientiousness, and stability come with (increasing) age, further evidence suggests high (average) openness is seen among the 20-24 age group, relative to other age groups. Suggesting younger Kenyans are more innovative. Extraversion is more prevalent among the 45-64 age group. Generally, age shows some useful relationship with all the personality traits.

Specifying non-cognitive Skills:

For use in this analysis, the personality traits are standardised. For ease of exposition, I will present one of the traits (Extraversion_av). Please see Appendix A2, for the rest of the traits. See also, the Data Appendix for more on the personality traits.

Notes– Source: Author's elaboration of the STEP Household Survey for urban Kenya. Please, see the Data Sections of Chapter 2 and related appendices.

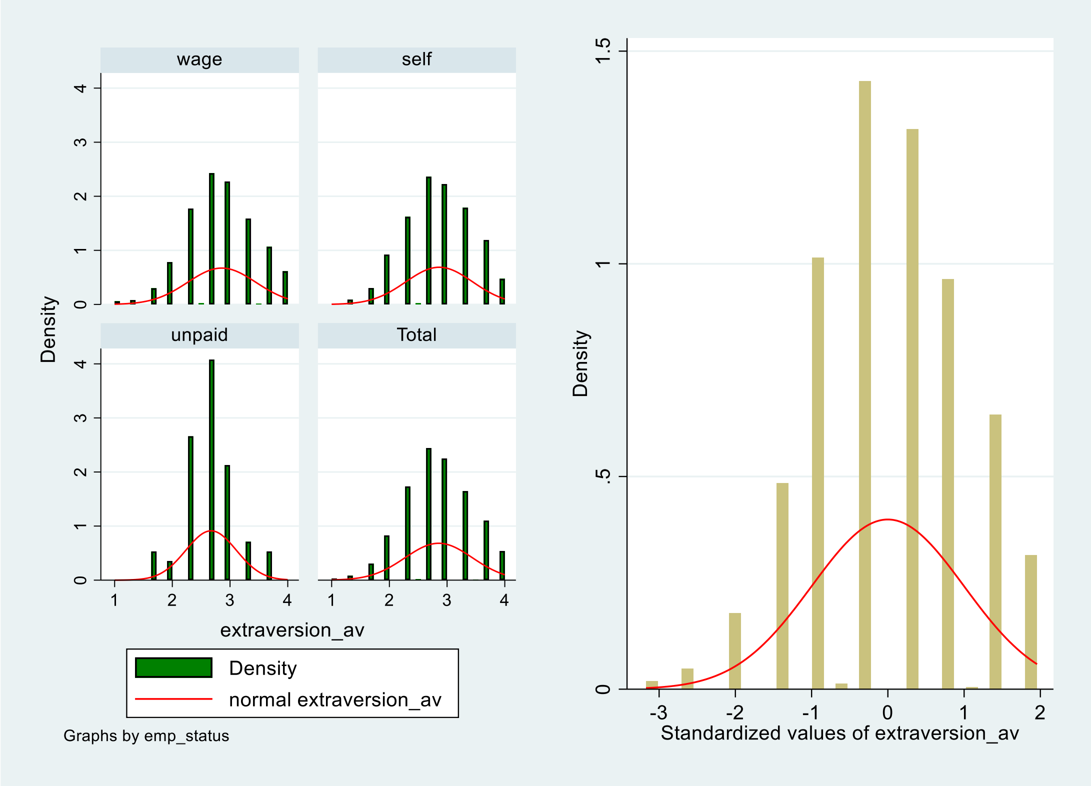


Figure 3-4 Extraversion Average, by Employment Categories.

Source: Author's elaboration of the STEP Household Survey for urban Kenya.

In module 6 of the STEP Household survey, the variable, *extraversion_av* (average of q01, q04, and q20) is used to capture the tendencies for positive emotions, being lively, active, and sociable. The STEP used a series of instruments to elicit this information from respondents (e.g., using a four-point frequency scale that ranges from *almost-always* to *almost-never*, are you talkative?). It is self-reported, as opposed to, a direct assessment of skill or trait. Hence, it is an indirect measurement.

Table 3-5 Survey Instruments for Extraversion_Average.

Trait	Questions in module 6 (G)	Items or survey instruments
Average of Extraversion	Q.1.01	Are you talkative?
	Q.1.04 *	Do you like to keep your opinions to yourself? Do you prefer to keep quiet when you have an opinion? *
	Q.1.20	Are you outgoing and sociable, for example, do you make friends very easily?

Source: The World Bank's STEP Household Survey for Kenya. Please, see the Data Section of the Appendix Chapter for more descriptive evidence for all the traits.

Where * is scaling reversal for negatively scored items. It is important to note that the *extraversion_av* which is a simple average of all the instruments used are continuous variables that range from 1 – 4, with 1 meaning almost-never which accentuates the lowest evidence of trait; 4 almost always which accentuates high evidence of trait; 2 and 3 are medium levels of the trait.

3.2.1.3 Further Descriptive Evidence for Earnings, Skill, and Schooling

Tables 3.6 and 3.7 present further descriptive evidence disaggregated across employment, and credential categories. These support arguments raised and related discussions in the results, and conclusions subsections.

Table 3-6 Descriptive Evidence, by Categories of the Employed and Genders.

	Mean						
	Percent (%)	Earnings (usd)	Apvlit_c	age	hours_d	years_educ_act	tenure
Total – Employed	100	3.733	174.108	29.540	8.666	10.320	54.430
Male	47.61	4.009	183.567	30.193	8.897	10.789	56.330
Female	52.39	3.366	165.510	28.947	8.368	9.893	51.970
Wage	56.43	3.502	182.249	30.892	8.703	11.153	51.936
Self	40.8	4.075	166.214	32.978	8.712	9.881	58.203
Unpaid	2.78	1.070	131.391	31.821	7.215	6.493	49.477
Private-Sector	88.15	3.195	176.877	30.663	8.755	10.763	48.926
Public-Sector	11.85	5.858	225.480	32.706	8.306	14.150	71.275
Entrepreneurship	17.13	8.906	185.763	33.566	8.866	11.970	66.863
Lone Employed	82.87	3.099	162.183	32.846	8.686	9.456	55.988

Notes— Source: Author’s elaboration of the STEP Household Survey for urban Kenya. Please, see Data Sections of Chapter 2 and related appendices of the Appendix Chapter.

Table 3.6 shows that the average number of years of schooling of the employed in urban Kenya is 10.32 years, this is higher than the average for sub-Sahara Africa (which is under 10 years) but below the average for OECDs (which can get as high as 18 years). The unpaid have the least schooling with an average of 6.5 years of schooling. On average those without qualification (ISCED01) have 1.63 years of schooling (Table 3.7). Across genders and employment categories (Table 3.6), men on average have a year of schooling more than women. Within the employed, the public sector wage-employed has the highest number of years of schooling, at 14.2 years.

Table 3-7 Descriptive Evidence, by Credential Categories.

	Mean						
ISCED - 97	Percent (%)	earnings_h_usd	apvlit_c	age	hours_d	years_educ_act	tenure
ISCED01	11.92	1.7088	60.324	34.030	8.268	1.630	64.850
ISCED1	22.72	2.120	137.883	29.445	8.871	7.906	55.199
ISCED2	14.06	3.111	191.273	28.303	8.696	10.151	55.975
ISCED34A	33.35	2.886	198.198	27.978	8.812	12.425	48.943
ISCED4B	8.64	5.311	230.668	30.128	8.721	14.585	54.029
ISCED56	9.31	9.934	243.443	30.945	8.230	16.100	56.269
Total	100	3.733	174.108	29.540	8.666	10.320	54.430

Notes—source: Author’s elaboration of the STEP Household Survey for urban Kenya. Please, see Data Sections of Chapter 2 and related appendices of the Appendix Chapter.

Interestingly from Table 3.7, the hourly earnings of the employed with ISCED2 are high relative to those of workers with higher educational attainment (workers with ISCED34A credentials). However, the ISCED2 workers have lower average skills (apvlit_c) compared to workers with the ISCED34A. With a relatively high mean hourly earnings and low average skill for ISCED2 compared to ISCED34A, with relatively low hourly earnings and high average skill (apvlit_c). This descriptive evidence appears to be confounding. However, examining the baseline (OLS) estimates of returns to both categories of the employed, in Table 3.8 suggests having the ISCED2 qualification (with an average of 10.2 years of schooling) explains a 44.3% rise in hourly earnings. However, the ISCED34A qualification (an average of 12.4 years of schooling) only explains a 42.2% rise in hourly earnings.

Acknowledging the substantial difference in schooling between these two groups, the employed with ISCED2 qualifications are (substantially) better remunerated. Interestingly, relative to those with the ISCED34A credential, those with the ISCED2 credential not only have higher tenure (months of work experience) but are also older and have lower supply, making up only 14.06% of the employed. However, those with ISCED34A make up 33.35% of the employed. Those with the ISCED34A credential have mainly attended TVET colleges. Respondents with ISCED2 have relatively low average reading proficiency scores compared to those with ISCED34A, evidence from Chapter 2 (efficiency in schooling, see Table 2.13) suggests an additional year of schooling with ISCED2 credential category substantially explains reading proficiency. Suggesting schooling (and possibly, employment) of those with ISCED2 are useful drivers of their productivity that explain their wage premium.

3.2.2 Empirical Framework

In this subsection, a discussion on the estimation and identification strategies presented relates to the analyses of this chapter (chapter 3) and the subsequent chapter (chapter 4). The identification strategy details the basis of the quasi-experimental approach used to draw causal inferences using the Two-Stage Least Squares (2SLS-IV) approach. In the estimation strategy, an adapted version of the Mincerian Wage Equation estimated by Ordinary Least Squares (OLS) provides the models underpinned by the Human Capital Framework, used for baseline estimations. Please see the details of the Theoretical Framework of the Mincerian Wage Equation in Appendix A1. I now turn to discuss the Estimation Strategy.

3.2.2.1 Estimation Strategy and the Endogeneity Problem

Estimation Strategy

The baseline analytical models adopted in this study are inspired by the model specified by Moretti (2006) and Liu (2007) that accounted for individual human capital in estimating the wage effects of aggregate human capital or the pecuniary external returns. Specifically, Liu (2007) derived his model by augmenting the measure of aggregate human capital at the district level to the Mincerian Wage Equation. However, in this study, drawing from their specifications, I present Equations 3.1 and 3.2 as thus:

I implement Equations 3-1 (parsimonious specification) and 3.2 (less parsimonious specification) to give a response to research question 1 (Q1)—estimates of the baseline return. I implement the Oaxaca-Blinder Decomposition methodology (see details in estimation strategy of Chapter 2) to decompose the wage differentials across subsamples of interest (employment and genders). I carry out a mediation analysis, by implementing variants of Equation 3.2 across all measures of the human capital of interest in this study (schooling, cognitive, and non-cognitive skills), assessing the mechanisms through which each measure of human capital impacts the outcome (earnings); and assessing their robustness.

Equation 3-1

$$\log(w_i) = \beta_0 + \rho_0 H_i + \beta_2 tenure + e_i$$

Equation 3-2

$$\log(w_i) = \beta_1 + \beta_2 H_i + \beta_X \boldsymbol{X_i} + \sigma_i$$

Equation 3-1 is the most parsimonious specification of the model used as a baseline of this study. $\log(w_i)$ is the log hourly earnings in USD; H_i is the measure of human capital (this includes the years of schooling (continuous), credentials, *isced* (categorical); for non-cognitive skills, I use the Big 5; for cognitive skills, I use adult reading proficiency. ρ_0 is the wage effect of the measures of human capital in the parsimonious specification of the baseline model. However, in the less parsimonious specification, β is the parameter of interest for private returns to measures of human capital. It is important to note that, besides estimating the effects of non-cognitive skills and schooling simultaneously (evidence from this specification suggests the most robust estimates of the effect of non-cognitive skills), other specification of the model only includes one dimension of the measures of human capital of interest (either, schooling, cognitive skill or non-cognitive skills). In 3.2, $\boldsymbol{X_i}$ is a vector that captures other variables of interest including typical variables controlled for, and β_X captures the wage effect of each of the variables in $\boldsymbol{X_i}$. Hence, the $\boldsymbol{X_i}$ of Equation 3-2 includes tenure as the number of months of experience, specified to have a linear relationship (differs from Liu (2007) in this manner) as this allows maximising the effects of other variables of interest, such as aggregate schooling that enters Equation 3.2 as a quadratic term.

Other individual characteristics such as gender and measure of health—BMI are accounted for in 3.2. For this chapter, the z_c[86] variables are covariates in dummies that capture district size (number of households across strata). Quite important for the next chapter is, $\boldsymbol{E_e}$, a measure of interest that captures aggregate schooling and skills across districts. Based on the specification of Moretti (2006) individual schooling, $H_i \;\; may\, corellate\, with\, \boldsymbol{E_e}$. Both $z_c \; and \;\; \boldsymbol{E_e}$ are derived from the **stratum variable**, please, see the next chapter for the detailed specifications of these variables. As with measures of human capital, I do not simultaneously account for $\boldsymbol{E_e}$ and z_c in the models, as doing so results in spurious estimates. β_0 and β_1 are the constant terms; $\sigma_i \; and \; e_i$ are the error terms in 3.1 and 3.2 respectively, accounting for aggregate human capital in 3.2 means σ_i captures much more, possible shocks that impact the demand and supply of human capital, unobservable components of human capital returns to unobserved skills in the district, and an error term, the transitory components of (individual) log of wages (see Moretti (2006)).

[86] Based on World Bank's STEP design, Kenya has four strata, hence, cities are categorised as thus: Nairobi; Cities with more than 100 000 Households; Cities with under 100 000 Households but above 60 000 Households; Lastly, other Cities, with under 60 000 Households.

Biases from Baseline Estimates

The mediation analysis and the descriptive evidence suggest that the specifications of Equation 3-2 provide estimates of non-cognitive skills on which causal statements can be made. However, neither estimates of cognitive skill and schooling from Equations 3-1 nor 3-2 are strong enough to make one believe the following is not the case ($E(H_{i.}, \sigma_i)$≠ 0 $E(H_{i.}, e_i)$≠ 0 $E(\boldsymbol{E_e}, \sigma_i) \neq 0$), although we may still interpret the function as a conditional expectation of $Iog\ w_i$ with H_i and $\boldsymbol{X_i}$ in place, but then $\beta_2\ OR\ \rho_0$ are not deemed causal effects of H_i on $Iog\ w_i$ due to endogeneity issues of H_i**.** Simply put, endogeneity of the H_i variable, entails a situation when a positive correlation occurs between H_i and other unobserved or otherwise omitted (ability) variables also affect earnings positively. If this happens, OLS will give estimates of returns that are biased upward, in other words, it overestimates the parameter, $\beta_2\ OR\ \rho_{0.}$. Besides, $H_{i.},\ estimates\ of\ \boldsymbol{E_e}$ **in** $\boldsymbol{X_i}$ deemed correlated to H_i may be biased.

The literature presents numerous studies documenting sources of biases and strategies for dealing with these biases. The literature presents some techniques explored to make causal statements of return estimates. Griliches (1977) was one of the early works that attempted to address endogeneity concerns by including an explicit measure of ability using IQ tests. This seems logical or plausible enough, as controlling for ability in models that estimate returns to schooling is an attempt to account for an unobserved factor that may strongly explain schooling, hence, accounting for such may give truer return estimates, however, the question is, what measure of ability is explicit enough? Or besides mere ability are other unobserved variables or factors that can strongly explain schooling? The work of Blackburn and Neumark (1993) showed that controlling for ability still presents an upward bias of return using Least Square estimates. As an extension to the first approach discussed, a somewhat non-conventional approach to address the endogeneity of schooling measure in the literature explored the use of siblings or twins with an understanding that levels of ability for siblings or twins are the same, hence it assumes no difference in ability and observes differences in earnings and schooling for twins/siblings. A major defect to this approach entails estimating returns to schooling with differences in schooling (between siblings or twins) as this is known to result in material measurement error that can lead to significant bias in return estimates. Blanchflower and Elias (1999), extracted twins' data from the National Child Development Survey panel for the UK, using test scores to control for ability. They came up with interesting findings on estimates of returns to schooling for twins using Least Squares even with evidence of an upward bias, they concluded, that returns estimates using twins were more plausible than those of non-twins. However, what is clear is that although the use of twins' data and the use of test scores as instruments for ability are not sufficient to objectively make causal claims or give estimates of returns to schooling that are causal to earnings, they are necessary in supporting causal statements, with further assumptions in place. This is at least, a step away from the mere association between schooling and earnings. One other method largely explored in the literature which sadly, cannot be implemented in this study due to data limitation is treating ability as a fixed effect using panel data (the dataset for this study is cross-sectional). It is important to note that using a panel dataset can also present issues in endogeneity, as return estimates may be obtained from respondents who return to school (Angrist and Newey, 1991) after a break. A widely used approach in attempting to achieve causality in return estimates entails an exogenous shock that impacts schooling, as an Instrumental Variable – this involves natural variation in the data– examples include, the season of birth, and proximity to educational institution. This is with the understanding that the season of birth and proximity to the site of the institution (educational) attended are exogenous in themselves, and they can influence schooling but not earnings or they can only influence earnings through schooling. Hence, these natural variations in the dataset are known to be due to exogenous changes in schooling. This approach has been explored extensively (Angrist Krueger and Card, 1992; Card, 1993), however, the evidence from the use of the natural experiment approach that uses plausible instruments has minimal explanation for the endogeneity concern of the measure of schooling (Harmon and Walker, 1995) and beyond that, such approach is known to result in material biases in estimates if not well specified. As a further extension to the use of 'natural variation in dataset', the approach explored in this study to obtain return estimates deemed to support causal statements on the effects of schooling on wages entails taking advantage of what is understood to be 'institutional aspects' of schooling as the basis of Instrumental Variables (IV). Examples of instruments with 'institutional aspects' of schooling include school expansion projects, the result of reforms; and changes in the compulsory school leaving age resultant of reforms. For reforms that impact schooling (and other measures of interest, in this case, measures of skills).

The study of Acemoglu and Angrist (2000) sets the scene for this study fully addressing biases to private and external (aggregate) returns to schooling, exploiting exogenous variations in schooling using measures of reform and quarters of birth. Inspired by the study of Acemoglu and Angrist (2000) in the United States of America, I address biases in the estimates of private and external returns in Kenya exploiting interactions of the 1985 curriculum reform[87] and Quarter-of-Birth that resulted in substantial variation in schooling and skill in Kenya.

The study of Acemoglu and Angrist (2000) set out the details of the 2SLS-IV (Two-Stage Least Square Instrumental Variables) approach to drawing causal inferences from estimates of external and private returns to schooling in urban Kenya, as thus:

Starting with the baseline model estimated with the OLS (see equation 3-2):

$\log(w) = \beta_1 + \beta_2 H + \beta_X \boldsymbol{X} + \sigma_i$ (let's take this as (1)) on modifying this, to reflect the aggregate schooling, we have this: $\log(w) = \beta_0 + \beta_1 H + \beta_{\boldsymbol{E_e}} \boldsymbol{E_e} + \sigma_e$ (let's take this as (2)). In (2) external and individual schooling are taken to be endogenous, however, in (1), only individual schooling is taken as endogenous.

Here, let us recall that the contention is $\text{cov}\,(\text{H}, \sigma_i) \neq 0$ and $\text{cov}\,(\boldsymbol{E_e}, \sigma_i) \neq 0$. Where both individual and private schooling are endogenous as in (2).

To start with the simplest case. Based on the simple 2SLS-IV approach by Angrist and Pischke (2009) that showed the useful basis to draw causal inference from estimates where only schooling is treated as endogenous—i.e., the typical IV studies for returns to schooling without aggregate schooling, will attempt to draw causal inference from the following using (1).

With a 'deemed' valid instrument for H such as a dummy for Quarters-of-Birth, as q* and the following conditions are met: $\text{cov}\,(\text{q}*, \text{H}) \neq 0$; and $\text{cov}\,(\text{q}*, \sigma_i) = 0$. Then using the 2SLS approach to estimate β_2 without X (controls), gives:

[87] As earlier discussed, schooling reforms in Kenya in 1985 resulted in the change from the 7-4-2-3 structure in place in 1963 to the current 8-4-4 that changed at the launch of the recent competency-based curriculum (CBC) in 2017 that entails 2-6-3-3-3 for pre-primary, primary, lower-secondary, upper-secondary, and tertiary schooling. The 1985 curriculum reform was motivated from arguments against the previous structure lacking flexibility and being too academic. The rationale for the change is deemed to have no direct association to earnings. The instrument validity or the relevance condition; and the exogeneity assumption. Simply put, the former entails, that the instrument should be correlated with the variable of interest, in this case, the measure of schooling; and the latter entails a requirement for the instrument to be uncorrelated with the error terms. See tests in A2 and first-stage equations in the Results and Discussion subsection.

$$\beta_2 = \frac{\text{cov}\,(w, \text{q}\,*)}{\text{cov}\,(\text{H}, \text{q}\,*)} = \frac{E[w\,|\text{q}\,*= 1] - E[w\,|\text{q}\,*= 0]}{E[H\,|\text{q}\,*= 1] - E[H|\text{q}\,*= 0]}$$

Then, β_2 may give local average treatment effects (LATE) or heterogenous treatment effects (HTE) as potential outcomes.

If H_1 denotes assignment to treatment; H_0 denotes assignment to control by p1985. As in, $H = \text{q} * H_1 + (1 - \text{q}\,*)H_0$. Angrist et.al (1996) made it clear that for defiers, $H_1 < H_0$ ($H_0 = 1$; $H_1 = 0$); for compliers, $H_1 > H_0$ ($H_0 = 0$; $H_1 = 1$); Never-takers, $H_1 = H_0 = 0$; Always-takers, $H_1 = H_0 = 1$. Although the inequality sign for compliers indicates experimental variation (or a perfect experiment where observations are rightly assigned to treatment and control). However, one cannot identify the groups individuals belong to as only a treatment indicator is observed, H_1 *or* H_0.

Hence with,

1. Independence: $(w_0, w_1, H_0, H_1) \perp\!\!\!\perp q\,*$

2. First stage equation such that, $0 < P(\text{q}\,*= 1) < 1; P(H_1 = 1) \neq P(H_0 = 1)$

3. Monotonicity, where $H_1 \geq H_0$

The above assumptions, identification is achieved (see Angrist et al., (1996); Imbens and Angrist (1994) for the full proof) as thus:

$$E\left[\,w_1 - w_0\,|H_1 > H_0\,\right] = \frac{\text{cov}\,(w, \text{q}\,*)}{\text{cov}\,(\text{H}, \text{q}\,*)} = \frac{E[w\,|\text{q}\,*= 1] - E[w\,|\text{q}\,* = 0]}{E[H\,|\text{q}\,* = 1] - E[H|\text{q}\,*= 0]}$$

Hence, the mean outcome $E\left[\,w_1 - w_0\,|H_1 > H_0\,\right]$ is the average treatment effect for compliers, deemed the Local Average Treatment Effect (LATE) which is the ATE for those impacted by their treatment (as opposed to their ‘control’) status. Hence, the LATE is strongly dependent on the instrument for H_1, $q\,*$ (in this case).

This (as below) is about the simplest case. However, Acemoglu and Angrist (2000) suggest external and private returns following (2).

However, with a valid instrument for $\boldsymbol{E_e}$ and not H such as the reform dummy (p1985). Where the following conditions are met: $\text{cov}\,(\text{p1985}, \boldsymbol{E_e}) \neq 0$; and $\text{cov}\,(\text{p1985}, \boldsymbol{E_e}) = 0$. Then using the 2SLS approach to estimate $\beta_{\boldsymbol{E_e}}$ without X (controls), gives:

$$\beta_{\boldsymbol{E_e}} = \frac{\text{cov}\,(w, \text{p1985})}{\text{cov}\,(\boldsymbol{E_e}, \text{p1985})} = \frac{E[w\,|\text{p1985} = 1] - E[w\,|\text{p1985} = 0]}{E\big[\boldsymbol{E_e}\,|\text{p1985} = 1\big] - E[\boldsymbol{E_e}|\text{p1985} = 0]}$$

Then, $\beta_{\boldsymbol{E_e}}$ can either gives local average treatment effects (LATE) or heterogenous treatment effects (HTE) as potential outcomes. However, we can see that the instrument, p1985 is valid for $\beta_{\boldsymbol{E_e}}$, suggesting $\beta_{\boldsymbol{E_e}}$ is causally identified. However, the instrument is invalid for H and yields no consistent estimate of the effects of H.

Acemoglu and Angrist (2000) also showed that, with external and private returns, following (2), certain adjustments (on subtracting the effects of human capital externalities) can make it possible for the instrument (p1985) to result in a consistent estimate of private returns. Which will result in a valid instrument for H and not $\boldsymbol{E_e}$ such that the following conditions are met: $\text{cov}\,(\text{p1985}, \text{H}) \neq 0$; and $\text{cov}\,(\text{p1985}, \sigma_i) = 0$. Then using the 2SLS approach to estimate β_2 without X (controls), gives:

$$\beta_2 = \frac{\text{cov}\,(w, \text{p1985})}{\text{cov}\,(\text{H}, \text{p1985})} = \frac{E\big[w - \beta_{\boldsymbol{E_e}}\boldsymbol{E_e}|\text{p1985} = 1\big] - E\Big[w - \beta_{\boldsymbol{E_e}}\boldsymbol{E_e}\,|\text{p1985} = 0\Big]}{E[H\,|\text{p1985} = 1] - E[H|\text{p1985} = 0]}$$

Then β_2 may give the local average treatment effects (LATE) or heterogenous treatment effects (HTE) as potential outcomes. In this case, $\beta_{\boldsymbol{E_e}}$ is not causally identified. However, based on (2), the individual and district-level schooling are treated as endogenous. Acemoglu and Angrist (2000) argue that the quarter of birth variable gives consistent estimates of private returns. Hence, they tested both (empirically), using several interactions of quarter of birth and year of birth; and compulsory schooling law instrument (synonymous to the reform dummy used) and found evidence of consistency in the estimates of private and external returns to schooling in estimating (2). In this study, the interactions of the reform dummy and Quarters-of-Birth (taken as p1985*) make it possible to meet the following conditions: $\text{cov}\,(\text{p1985}*, \text{H}) \neq 0$; and $\text{cov}\,(\text{p1985}*, \sigma_i) = 0$. Hence, the p1985* may be a valid instrument for both $\boldsymbol{E_e}$ and H that can be estimated simultaneously, using the 2SLS-IV approach. Evidence from the first-stage equation is presented.

It is important to note that, relative to return estimates from OLS, return estimates correcting for endogeneity through IV tend to be significantly higher in value and deemed robust to changes in specifications, especially with changes in the instruments. However, a misspecified IV procedure may lead to return estimates that may be misleading or (much) larger than OLS estimates. However, to avoid misspecification of the IV procedure, a good instrument for the IV procedure must be tested beyond meeting the conditions previously discussed. I argue that the return estimate using the interactions of the reform dummy and the quarters of birth (Q2, Q3 & Q4) as instruments for the procedure provide the average marginal returns to schooling and skills, following the works of Card (2001), specifically, for those affected by the reform– whose years of schooling changed due to the reform – this explains the LATE—Local Average Treatment Effects. Hence, useful, or sufficiently robust estimates are deemed to provide a suitable basis for the causal statement, being well-specified. Returns using the IV procedure of the 2SLS approach account for other sources of bias, including those resulting from measurement errors of the measure of schooling. I now turn to specify the models that aid in operationalising the identification strategy. I also exploit this approach, using individual skill as the outcome variable and individual and aggregate schooling as predictors (see Chapter 2). Findings suggest that schooling explains skills, further suggesting that, the effect of schooling on earnings is through skills, in support of the human capital theory.

3.2.2.2 Identification Strategy

To obtain estimates of the reform-affected returns (Q2 – Research Question 2) to schooling and skill, I implement Equation 3-4, instrumenting schooling, and skill with the interaction of the reform dummy (p_{1985}) and Quarters of Birth as Instrumental Variables (IV) using a parsimonious specification of the Wage Equation. See Appendix A3 For detailed tests for the endogeneity of schooling and skill and tests for the instruments. I explore the use of the Two-Stage Least Square Approach as thus:

Equation 3.3

Where,

$\boldsymbol{E_e} + H_\ell = \beta_{1985} p_{1985}\#\text{Q2} + \beta_B p_{1985}\#\text{Q3} + \beta_X p_{1985}\#\text{Q4}$ as the first stage equation.

Equation 3.4

$$\log(w_i) = \beta_\beta + \rho\ (\boldsymbol{E_e} + H_\ell = \beta_{1985} p_{1985}\#\text{Q2} + \beta_B p_{1985}\#\text{Q3} + \beta_X p_{1985}\#\text{Q4}) + \alpha_i$$

Here, I argue that ρ supports a causal statement on the wage effects of H_{ℓ} (individual schooling and skill) and $\boldsymbol{E_e}$ (aaggregate schooling and skills). Again, all variables are as previously described under Equations 3-3 with β_{β} as the constant and α_{i} as the error term.

3.3 Results and Discussion

In examining the private returns to education and skills as measures of human capital. Although accounted for in the models for private returns to education and skill, the human capital externalities are examined in the subsequent chapter. However, preliminary findings from this chapter suggest that although the externality of skill is positive and enters linearly, the externality of education (years of schooling) enters as a quadratic term, and it is negative but increasing (becoming less negative) with an increase in average schooling across cities. Accounted for across most specifications of the models are dummy variables that account for the number of households in districts across strata, using the dummy for Small Cities[88] as the reference category. Overall, the findings suggest, the strata- effects adversely impact hourly earnings in urban Kenya. Particularly, relative to Small Cities, the effects are most adverse in Other Large Cities, then Mid-Sized Cities then Nairobi, in this order. I now discuss the main findings of this chapter devoted to examining the private returns to education and skills as measures of human capital[89]. However, in addition to examining the private returns to education and skill. I emphasise the effects of time-varying covariates such as age and experience, and other key variables accounted for across subsamples of interest, these include, gender, effects of non-cognitive skills and Body Mass Index (BMI) as

[88] As earlier highlighted (see Data Section), the STEP sampling follows a three-stage stratified sampling where respondents are drawn from districts across the four different strata created based on the number of households. These include the Nairobi stratum; Other Large Cities stratum, having more than 100,000HHs; Mid-Sized Cities stratum, with over 60 000HHs; Small Cities stratum with under 60 000HHs.

[89] In examining private returns to education and skill, I do the following: This entails an examination of the effects of the skills and education of the individual on their wage. To do this, I start with schooling. I present and discuss main results showing different specification of the model (controlling for several other variables) effects of both continuous schooling (years_educ_act) and categorical (isced) schooling. The former presents return estimates for an additional year of schooling regardless of credential categories and the latter presents return estimates of each of the credential categories. Next, I examine return estimates across genders and employment categories, specifically, informality, wage, and self-employment. I examine how differences in characteristics (particularly, schooling) and potential discrimination explain wage gaps across these groups. Finally, I turn to quasi-experiments to obtain more consistent estimates of returns to education. I then turn to examine the wage effects of cognitive skills as a measure of human capital, I carry out similar analyses as the former (schooling). In concluding this analysis, I compare effects of schooling and skills as measure of human capital examining arguments raised by Pritchett (2001) and Hanushek (2015) that suggest mere schooling may not reliably capture human capital (in urban Kenya) in developing context as they argued schooling gives minimal or no useful skills for productivity and growth.

discussed in the introduction[90]. Across the main analytical sample (see Table 3.8), age and months of work experience (tenure) enter (the model) as linear terms and have positive effects on hourly earnings. Evidence suggests that a year increase in age explains a rise in hourly earnings ranging from 0.61% to 0.79%, statistically significant at the 1% level at best. However, across subsamples of the main analytical (see Table 3.9), evidence suggests that an increase in age only explains earnings for the female gender and the private-sector wage employed where return to an additional year of age explains a rise in hourly earnings in the range: (1-1.4) %. This is consistent across the model specifications accounting for continuous and categorical schooling. Similarly, across the model specifications of the main analytical sample (Table 3.8), an additional month of work experience accounts for an average rise in hourly earnings ranging from 0.15% to 0.23%. However, across subsamples of the main analytical sample (see Table 3.9), the evidence suggests that only the male gender; the informally and private-sector-wage employed; and those in entrepreneurship have statistically significant wage effects on their monthly experience and this range from 0.14% to 0.43%, mean coefficients are substantially (relatively) great for those in entrepreneurship. Although with a modest mean coefficient, the wage effect of work experience is strongly statistically significant, particularly, for the private-sector wage employed where age and work experience are statistically significant at 1% level at the very least. Interestingly, both time-varying covariates only enter the model linearly, suggesting work experience strongly impacts earnings, particularly, for the economically active in the labour market of urban Kenya. Turning to the effects of BMI and non-cognitive skills, across the model specifications of the main analytical sample (Table 3.8). Evidence suggests, the mean effect of BMI is low, ranging from 1.33% to 1.39% statistically significant at the 5% level, however further evidence from subsamples (Table 3.9) suggests BMI only favourably impacts the hourly earnings of males and the formally employed. The wage effects on hourly earnings range from 1.65%-1.93%. Across all the five personality traits considered in this analysis, evidence from the main analytical sample (Table 3.8) suggests, whilst Extraversion weakly explains earnings with a mean effect of a rise of about 7.5% in hourly wage, the 5% statistical significance is lost on accounting for average schooling in lieu of strata-specific effect. However, Openness to Experience best explain hourly earnings with mean coefficient ranging from 11% to 17% rise in hourly earnings. Effects are strongest in models that include continuous schooling. However, heterogeneity analysis (Table 3.9) unpacks useful wage

[90] The former (non-cognitive skill) gives some insights into the wage effects of useful personality traits or abilities and the latter (BMI) gives insights into the effects of health conditions on hourly earnings. Understanding the effects of these two useful measures of human capital improve understanding of human capital in urban Kenya providing useful estimates that can be compared to similar estimates in developing contexts.

effects of non-cognitive skills across subsamples of interest. Evidence suggests only the openness to experience has consistent wage effects on the hourly earnings of the male gender, the informally and the private-sector wage employed with effects strongest for the male gender with mean coefficients ranging: (19.3-24.5) %, statistically significant at the 0.1% level. Further evidence suggests no consistency in the wage effects of Extraversion. Interestingly, evidence suggests the wage effect of conscientiousness is consistent for females, the informal and lone self-employed, suggesting that the mean coefficients are greatest for the lone employed however, effects are strongest (at the 1% level of statistical significance) for the female gender. Hence, conscientiousness in females raise hourly earnings by the range: (21.3-23) %. Finally, being agreeable raises hourly earnings by the rage: (14.8-15.2) % of the formally employed. Table 3.8 shows model specifications have between 21% and 29% explanatory power of hourly wage, with models including continuous schooling explaining lower variation in hourly earnings, relative to model specifications that include categorical schooling.

Table 3.8 reports the baseline (OLS) estimates of the private returns to education of the main analytical sample, accounting for non-cognitive skills and BMI, a measure of experience, age, gender, strata-specific effects, and externality of schooling as controls as earlier discussed in the previous subsection. Similarly, Table 3.9 presents similar specifications of columns (13) and (14) of Table 3.8 (accounting for the externality of schooling instead of strata-specific effects across subsamples of the main analytical sample.

3.3.1 What is the (Private) Return to Schooling in Urban Kenya?

The private returns to education are comparable across all specifications of the (OLS) model. To start with, without accounting for levels of schooling (credentials), the return to an additional year of schooling ranges from 10.2% to 10.9%. Turning to categorical schooling, overall, estimates of returns to schooling rise with the level of education (across credential categories) which is consistent with previous (and relatively recent studies—see Kimenyi et al., 2006; Wambugu, 2003; and Appleton et al., 1999) findings for Kenya with least returns at low credential categories and greatest returns at the highest credential categories. This suggests that it is now most beneficial for individuals to obtain tertiary education as returns are greatest at this level of schooling relative to other credential categories. However, whilst previous studies find positive and statistically significant returns to primary education, findings from this study suggest returns to mere completion of primary education have no statistically significant effects in urban Kenya. Suggesting, mere completion of primary education has a return that is not different from nil. This indicates a material (substantial) change in the trend in returns to schooling over time[91].

Particularly, the early work of Thias and Carnoy (1972) in Kenya finds the OLS returns to primary schooling to be 32.7% for males (and only 9.5% for females) and that of tertiary education to be only 27.4% for males. This change in trend is (mainly) attributable to changes in the supply of education; and the nature of skills demanded, over time. To obtain estimates on which causal inferences are drawn, Table 3.14 presents the output of the Two-Stage Least Square Instrumental Variables (2SLS-IV) Approach. Evidence suggests the return to an additional year of schooling is not statistically significant and, hence, not different from nil. This finding is for the pool (all) and does not account for heterogeneity across employment and credential categories. I now turn to examine return estimates and possible wage differentials across subsamples (of interest) of the pool.

[91] The study of Psacharopoulos (1978) suggests, relative to other categories of schooling, the returns (private and social) to primary schooling are (substantially) high, particularly for most of Africa, relative to other regions of the world. This was a basis for the substantial investment (or educational expansion) in schools, particularly in primary education in developing contexts.

Table 3-8 OLS Estimates, Returns to Education: Baseline Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	ln_earnings_h_usd													
years_educ_act	0.104***		0.109***		0.105***		0.109***		0.104***		0.105***		0.102***	
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
isced1		-0.024		-0.013		-0.018		-0.008		-0.011		0.008		0.012
		(0.721)		(0.847)		(0.777)		(0.912)		(0.870)		(0.908)		(0.855)
isced2		0.443***		0.390***		0.429***		0.389***		0.427***		0.353**		0.398***
		(0.000)		(0.000)		(0.000)		(0.001)		(0.000)		(0.002)		(0.000)
isced34A		0.422***		0.474***		0.445***		0.474***		0.449***		0.456***		0.444***
		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)
isced4B		1.050***		1.077***		1.052***		1.076***		1.052***		1.004***		0.994***
		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)
isced56		1.691***		1.723***		1.693***		1.714***		1.683***		1.667***		1.650***
		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)
Tenure	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age							0.006*	0.004	0.008**	0.005	0.006*	0.0037	0.008**	0.005
							(0.028)	(0.179)	(0.003)	(0.053)	(0.024)	(0.170)	(0.004)	(0.057)
Gender							-0.104*	-0.088	-0.112*	-0.093	-0.106*	-0.098*	-0.115*	-0.102*
							(0.036)	(0.071)	(0.026)	(0.057)	(0.036)	(0.047)	(0.026)	(0.039)
Bmi											0.013*	0.014*	0.014*	0.014*
											(0.035)	(0.028)	(0.028)	(0.028)
extraversion_av											0.077*	0.076*	0.046	0.052
											(0.046)	(0.027)	(0.237)	(0.132)
conscientiousness_avg											0.083	0.080	0.0870	0.085
											(0.167)	(0.138)	(0.137)	(0.106)
openness_av											0.170***	0.109**	0.164***	0.108**
											(0.000)	(0.002)	(0.000)	(0.003)
stability_av											0.031	0.050	0.033	0.054
											(0.560)	(0.332)	(0.536)	(0.313)
agreeableness_av											-0.035	-0.023	-0.028	-0.017
											(0.352)	(0.497)	(0.462)	(0.628)
stratum_N			-0.161*	-0.072			-0.129	-0.051			-0.127	-0.060		
			(0.030)	(0.200)			(0.084)	(0.372)			(0.089)	(0.309)		
stratum_L			-0.386***	-0.299***			-0.354***	-0.278***			-0.358***	-0.292***		
			(0.000)	(0.000)			(0.000)	(0.000)			(0.000)	(0.000)		
stratum_M			-0.269***	-0.227***			-0.239***	-0.209***			-0.245***	-0.226***		
			(0.000)	(0.000)			(0.000)	(0.001)			(0.000)	(0.000)		
avg_yos_					-0.430***	-0.247**			-0.396***	-0.227**			-0.403***	-0.251**
					(0.000)	(0.004)			(0.000)	(0.007)			(0.000)	(0.002)

_avg_yos_sq_					0.021***	0.012**			0.020***	0.011**			0.020***	0.012**
					(0.000)	(0.003)			(0.000)	(0.004)			(0.000)	(0.001)
_cons	-0.644***	-0.029	-0.376***	0.244**	1.576**	1.306**	-0.522**	0.167	1.199*	1.081*	-1.786***	-1.003***	0.041	0.063
	(0.000)	(0.638)	(0.000)	(0.003)	(0.002)	(0.005)	(0.002)	(0.185)	(0.021)	(0.026)	(0.000)	(0.001)	(0.941)	(0.901)
N	2224	2224	2224	2224	2224	2224	2224	2224	2224	2224	2114	2114	2114	2114
R-sq	0.176	0.253	0.194	0.266	0.185	0.257	0.199	0.268	0.191	0.260	0.220	0.285	0.213	0.277
adj. R-sq	0.175	0.251	0.192	0.262	0.183	0.254	0.196	0.264	0.189	0.257	0.215	0.279	0.208	0.271

Note: Table reports outputs of some variants of equations 3.3 and 3.4, as columns 1-14 with the outcome as log hourly earnings in USD. The predictors of interest are the continuous measure of schooling, years_educ_act, and categorical measures of schooling. Controls include strata-specific effects (based on the number of households across cities) and include, Nairobi, strata_N, other large cities, strata_L with over 100 000HH, medium cities, strata_M with over 60 000HH but under 100 000HH, and other cities under, strata_S with under 60 000HH (reference category). Other control variables include the average schooling, avg_yos, across districts, which also enters as a quadratic term. Other covariates include age, an indicator of female gender, tenure or number of months in current employment, BMI, and the measures of non-cognitive skills as the Big 5. p-values in parentheses: * p<0.05; ** p<0.01; *** p<0.001. With clustered robust standard error, at the district level.

Table 3.9 makes the first stage of the decomposition analysis that gives the OLS estimates of return, and useful insights into the basis (differences in characteristics (inequality) or potential discrimination (inequity)) of the wage gap across subsamples of interest in this study. Evidence from the outputs indicates substantial variations in the explanatory powers of the model across subsamples of interest. The model explains over thirty (30%) of the variation in wage for those in entrepreneurship and for those in public-sector wage-employment; however, only about 12% variation in the hourly wage of the informally employed and 20% variation in hourly earnings for the female gender are explained by the model. Overall, consistent with the output of the main analytical sample, model specifications that include categorical schooling have (substantially) high coefficient of determination relative to those with continuous schooling. This suggests the specifications that include categorical schooling are more robust, relative to specifications that include continuous schooling.

Table 3-9 OLS Return Estimates, by Genders and Employment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	Log Hourly Earnings in USD																	
	Pool		Gender (Female (1)/Male (0))				Informal (1)/Formal (0)				Wage-Employed (Public-Sector (1)/Private-Sector (0))				Self-Employed (Lone (1)/Entrepreneurship (0))			
	Continuous	Categorical	1		0		1		0		1		0		1		0	
Yos	0.103***		0.106***		0.099***		0.069***		0.159***		0.143***		0.107***		0.058***		0.087**	
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.010)	
isced1		0.011		0.132		-0.085		0.022		-0.444		1.486		0.090		-0.123		-0.041
		(0.863)		(0.236)		(0.266)		(0.758)		(0.088)		(0.110)		(0.256)		(0.259)		(0.916)
isced2		0.408***		0.806***		0.122		0.247*		0.574*		1.706		0.493***		0.212		-0.360
		(0.000)		(0.000)		(0.399)		(0.032)		(0.011)		(0.087)		(0.000)		(0.094)		(0.555)
isced34A		0.455***		0.486***		0.384***		0.353***		0.284		1.336		0.474***		0.243		0.098
		(0.000)		(0.000)		(0.000)		(0.000)		(0.217)		(0.161)		(0.000)		(0.050)		(0.799)
isced4B		1.005***		1.253***		0.810***		0.852***		0.759**		1.731		1.091***		0.699**		0.439
		(0.000)		(0.000)		(0.000)		(0.000)		(0.002)		(0.083)		(0.000)		(0.003)		(0.219)
isced56		1.665***		1.841***		1.491***		1.510***		1.287***		2.145*		1.612***		1.675***		1.291**
		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.035)		(0.000)		(0.000)		(0.003)
Tenure	0.002***	0.002***	0.001	0.001	0.002**	0.002**	0.001**	0.001**	0.001	0.001	0.0004	-0.001	0.002**	0.002**	0.001	0.001	0.004*	0.004**
	(0.000)	(0.000)	(0.068)	(0.133)	(0.003)	(0.004)	(0.009)	(0.006)	(0.153)	(0.102)	(0.752)	(0.648)	(0.003)	(0.002)	(0.471)	(0.376)	(0.029)	(0.008)
Age	0.008**	0.005*	0.014*	0.013**	0.004	0.002	0.007*	0.004	0.009	0.010	0.004	0.009	0.014***	0.010**	-0.002	-0.003	0.016	0.008
	(0.003)	(0.049)	(0.012)	(0.009)	(0.199)	(0.588)	(0.032)	(0.196)	(0.165)	(0.147)	(0.734)	(0.495)	(0.000)	(0.002)	(0.672)	(0.411)	(0.169)	(0.490)
Bmi	0.013*	0.013*	0.011	0.010	0.018*	0.019*	0.012	0.012	0.017*	0.019**	0.026	0.031	-0.0004	0.002	0.009	0.008	0.027	0.020
	(0.042)	(0.040)	(0.179)	(0.172)	(0.017)	(0.018)	(0.165)	(0.120)	(0.012)	(0.005)	(0.161)	(0.063)	(0.955)	(0.763)	(0.390)	(0.416)	(0.139)	(0.240)
extra_av	0.045	0.051	0.043	0.006	0.048	0.083*	0.030	0.043	0.023	0.072	-0.083	-0.026	0.029	0.046	0.018	0.042	0.310*	0.286
	(0.256)	(0.146)	(0.467)	(0.915)	(0.265)	(0.048)	(0.433)	(0.278)	(0.676)	(0.185)	(0.412)	(0.791)	(0.520)	(0.299)	(0.732)	(0.427)	(0.031)	(0.070)
cons_avg	0.088	0.085	0.230**	0.213**	-0.027	-0.029	0.156*	0.140*	-0.100	-0.136	-0.011	-0.162	-0.019	-0.016	0.243*	0.242*	0.110	0.151
	(0.139)	(0.108)	(0.008)	(0.007)	(0.659)	(0.639)	(0.031)	(0.042)	(0.271)	(0.108)	(0.937)	(0.249)	(0.726)	(0.735)	(0.028)	(0.024)	(0.683)	(0.529)
open_av	0.171***	0.113**	0.0804	0.026	0.245***	0.193***	0.165**	0.116*	0.137	0.091	0.103	0.042	0.215***	0.153**	0.139	0.096	0.047	0.140
	(0.000)	(0.002)	(0.128)	(0.626)	(0.000)	(0.000)	(0.003)	(0.023)	(0.059)	(0.183)	(0.398)	(0.764)	(0.000)	(0.002)	(0.083)	(0.183)	(0.850)	(0.560)
stab_av	0.041	0.061	-0.011	0.038	0.054	0.048	0.035	0.069	0.034	0.005	-0.174*	-0.101	0.045	0.065	0.124	0.144	-0.157	-0.222
	(0.436)	(0.248)	(0.879)	(0.582)	(0.374)	(0.445)	(0.525)	(0.214)	(0.674)	(0.953)	(0.042)	(0.222)	(0.434)	(0.248)	(0.235)	(0.140)	(0.381)	(0.241)
agree_av	-0.031	-0.020	-0.027	-0.022	-0.039	-0.023	-0.090*	-0.074	0.148*	0.152*	0.108	0.240*	-0.040	-0.032	-0.042	-0.026	0.040	-0.052
	(0.406)	(0.568)	(0.636)	(0.707)	(0.442)	(0.619)	(0.042)	(0.097)	(0.047)	(0.028)	(0.332)	(0.017)	(0.426)	(0.495)	(0.585)	(0.719)	(0.812)	(0.741)
avg_yos	-0.410***	-0.256**	-0.455**	-0.271	-0.356**	-0.246*	-0.363**	-0.269*	-0.502***	-0.243	-0.642	-0.412	-0.506***	-0.343***	-0.315	-0.198	-0.327	-0.411
	(0.000)	(0.002)	(0.005)	(0.090)	(0.002)	(0.036)	(0.002)	(0.018)	(0.000)	(0.073)	(0.069)	(0.303)	(0.000)	(0.001)	(0.103)	(0.313)	(0.591)	(0.509)
avg_yos2	0.020***	0.013**	0.023**	0.014	0.018**	0.012*	0.019**	0.014*	0.021***	0.011*	0.027	0.018	0.024***	0.016***	0.017	0.012	0.017	0.020
	(0.000)	(0.002)	(0.006)	(0.080)	(0.002)	(0.042)	(0.003)	(0.021)	(0.000)	(0.049)	(0.074)	(0.280)	(0.000)	(0.001)	(0.093)	(0.279)	(0.527)	(0.451)
_cons	0.001	0.028	-0.138	-0.362	-0.018	0.236	0.027	0.066	0.652	0.644	2.493	1.028	0.982	0.855	-0.366	-0.559	-0.745	1.002
	(0.999)	(0.956)	(0.884)	(0.702)	(0.978)	(0.701)	(0.968)	(0.918)	(0.322)	(0.436)	(0.213)	(0.699)	(0.158)	(0.175)	(0.730)	(0.602)	(0.840)	(0.796)

N	2114	2114	897	897	1217	1217	1553	1553	560	560	154	154	1084	1084	715	715	148	148
R-sq	0.211	0.275	0.197	0.280	0.225	0.287	0.119	0.167	0.298	0.336	0.322	0.351	0.266	0.320	0.087	0.157	0.238	0.306
adj. R-sq	0.206	0.270	0.186	0.267	0.218	0.085	0.112	0.159	0.283	0.316	0.265	0.275	0.258	0.310	0.071	0.137	0.170	0.221

Note: Table reports outputs of a variant of Equation 3.4, as columns 1-18 for subsamples of interest (Gender (3-6); Informality of Employment (7-10); Wage-Employment (11-14); Self-Employment (15-18)). The outcome variable is the log hourly earnings in USD. The predictors of interest are the continuous measure of schooling, years_educ_act, and the categorical measure of schooling is *isced*. Control variables include the average schooling, avg_yos, across districts, which also enters as a quadratic term. Other covariates include age, tenure or number of months in current employment, BMI, and the measures of non-cognitive skills as the Big 5. p-values in parentheses: * p<0.05; ** p<0.01; *** p<0.001. With clustered robust standard error, at the district level.

Baseline (OLS) Return Estimates, Explaining the Wage Gap, and Causal Effects of Schooling for Sub-samples of the Analytical Sample.

Genders in urban Kenya

Recent evidence of the OLS returns to schooling for females are greater than those of males. This is consistent with evidence from this study (this is consistent for both continuous and categorical schooling, see columns (3)-(6) of Table 3.9). Specifically, regardless of credential categories (using continuous schooling), the returns to an additional year of schooling are 10.6% for females and 9.9% for males. Both estimates are statistically significant at the 0.1% level. However, accounting for educational attainment (using categorical schooling), the evidence suggests that, although post-primary schooling (isced2) has statistically insignificant returns for males, females have a return for their post-primary schooling that is greater than (their) secondary (and some post-secondary) schooling (isced34A). The findings from this study are consistent with the (relatively) recent work of Kimenyi (2006) that also accounted for externalities of schooling, in Kenya and found returns to continuous and categorical schooling greater for the female gender, relative to the male gender. However, this is inconsistent with past studies where OLS returns to schooling for males are substantially greater than those of the female gender (see Thias and Carnoy, 1969; and Manda et al., 2002). As earlier argued, this suggests the supply, demand, and nature of skill across (particularly for females with post-primary) credential categories may be driving such returns. Moreover, Table 3.8 suggests that being a woman is associated with a loss of between 10% to 12% of hourly earnings, statistically significant at the 5% level (see columns 11-14). I now turn to examine the wage gap between genders.

Table 3.10 decomposes the wage differential across genders using the Oaxaca-Blinder Decomposition method. Although the specified (fitted) model explains a 68.8% rise in hourly earnings for the male gender however it (the fitted model) only explains a 45.4% rise in hourly earnings for the female gender, resulting in a wage gap (or differential in wage) of 23.4% between both genders. This is a shortfall for the female gender relative to the male gender. Further evidence suggests about 51% of the wage differential is attributable to differences in characteristics (particularly, schooling alone accounts for about 74% (8.76/11.9) of these characteristics) and about 49% of the wage differential is attributable to potential discrimination (this is due to return effects of non-cognitive skills). Therefore, although the differences in schooling characteristics alone account for over 37% (8.76/23.4) of the wage gap across genders; return effects of personality traits, particularly, potential discriminations based on the reward for conscientiousness and openness to experience across genders, account for the wage gap. This suggests that relative to the males, the females are rewarded for their hard work (or conscientiousness) in the market; and relative to females, males enjoy a premium for their openness to experience in the labour market. To obtain more consistent estimates of returns to schooling across genders (see Table 3.14), the evidence from the quasi-experimental approach shows that, the (causal) effect of an additional year of schooling is not different from nil for the female gender and the return to an additional year of schooling for the male gender is 20%. This is statistically significant at the 1% level. I now discuss the wage gap and estimates of returns to education across the formal and the informally employed.

Table 3-10 Decomposition of Wage Differential, by Genders

		Differential											
		Prediction 1 (Male)		0.688***									
				(0.000)									
		Prediction 2 (Female)		0.454***									
				(0.000)									
		Difference		0.234***									
				(0.000)									
	Total	YOS	Tenure	Age	BMI	Extra	Cons	Open	Sta	Agree	A_YOS	A_Yos_2	
Endowment	0.119***	0.088***	0.007	0.009	-0.004	0.0003	0.008	0.007	-0.001	0.0003	-0.014	0.018	
	(0.000)	(0.000)	(0.141)	(0.184)	(0.267)	(0.839)	(0.184)	(0.164)	(0.877)	(0.762)	(0.750)	(0.691)	
Return	0.115*	-0.072	0.014	-0.312	0.186	0.014	-0.832**	0.485*	0.172	-0.036	1.025	-0.485	0.120
	(0.025)	(0.577)	(0.786)	(0.145)	(0.430)	(0.935)	(0.006)	(0.037)	(0.386)	(0.870)	(0.627)	(0.665)	(0.915)
Interaction	0.001	-0.006	0.001	-0.006	-0.003	0.00003	-0.009	0.015	0.005	0.0001	0.003	-0.004	
	(0.954)	(0.580)	(0.787)	(0.289)	(0.466)	(0.940)	(0.184)	(0.080)	(0.397)	(0.880)	(0.789)	(0.768)	

Note: The table reports the wage differential and decomposition of the differential across genders (see first-stage equations in Table 3.9) showing the effects of characteristics, return, and the interaction of characteristics and returns that explain the differential in wage across genders. Accounting for key variables, as in Table 3.9. The p-values in parentheses: * $p<0.05$; ** $p<0.01$; *** $p<0.001$. With clustered robust standard error, at the district level.

The Formal/Informal in Urban Kenya

Evidence from Table 3.9 suggests that the OLS returns to an additional year of schooling for the informally employed is substantially less than (under half of) that of the formally employed. For the informal, the return is 6.9% and for the formal, it is 15.9% using continuous schooling. However, across credential categories, evidence suggests, it is more remunerative to be informally employed relative to being formally employed, this is the case for all from those with secondary and post-secondary credentials to those with tertiary education. It is only more beneficial to be employed formally, with post-primary schooling (isced), interestingly, having secondary and post-secondary schooling (isced34A) explains nil earnings for the formally employed. Suggesting the type of skills acquired in secondary and some post-secondary schooling is only relevant in informal employment, mainly made of the self-employed. Almost eighty (80) per cent of the workforce in urban Kenya is informally employed (see descriptive evidence). Understanding the basis of the wage gap across the groups (or employment categories) of the workforce unravels useful growth (policy) insights. Evidence from the fitted model presented in Table 3.11 suggests a substantial wage rise for the formally employed (this order is well over three times) relative to the informally employed. The wage gap between the formal and the informally employed is an 87% rise in hourly earnings for the formally employed relative to the informally employed (which also presents a shortfall of 87% in the hourly earnings of the informal relative to the formal). This wage gap is due to differences in characteristics, potential discrimination, and the interaction of the differences in characteristics and potential discrimination. Schooling strongly influences the differences in characteristics, returns (or potential discrimination), and the interaction of the characteristic and return effects

One would expect differences in characteristics, particularly schooling characteristics to explain the wage gap between the formally and informally employed, however, evidence suggests a statically significant difference in the openness also explains the wage gap for these groups of the employed. Interestingly, about thirty-three percent (33) % of this wage gap is attributable to potential discrimination. Further evidence suggests this is due to the discrimination in the schooling of the informally employed, relative to the schooling of the formally employed. This suggests a case where (even with homogeneity in schooling and skills), the informal is less paid relatively. Further evidence suggests that the returns to the conscientiousness and agreeableness of the informally and formally employed respectively are sources of potential discrimination between the two groups. Again, this is indicative of a case where relative to the formally employed, the informally employed are less rewarded for their conscientiousness; and relative to the informally employed, the formally employed are more rewarded for their agreeableness. To obtain more consistent estimates of returns to schooling across gender (see Table 3.14), whilst evidence from the quasi-experimental approach shows that the (causal) effect of an additional year of schooling is not different from nil for both the informal and the formal. I now turn to examine the returns to schooling and the wage gap across the wage-employed emphasising the public sector and private-sector wage-employed.

Table 3-11 Decomposition of Wage Differential, by Formal/Informal

		Differential											
		Prediction 1 (Formal)		1.228***									
				(0.000)									
		Prediction 2 (Informal)		0.359***									
				(0.000)									
		Difference		0.870***									
				(0.000)									
	Total	YOS	Tenure	Age	BMI	Extra	Cons_avg	Open	Sta_av	Agree	A_YOS	A_YOS^2	
Endowment	0.320***	0.249***	0.011	0.002	0.001	0.003	0.010	0.028**	0.002	-0.004	-0.193*	0.211*	
	(0.000)	(0.000)	(0.086)	(0.484)	(0.663)	(0.445)	(0.159)	(0.006)	(0.569)	(0.225)	(0.016)	(0.021)	
Return	0.284***	0.884***	-0.012	0.074	0.124	-0.019	-0.827*	-0.081	-0.002	0.681**	-1.417	0.246	0.625
	(0.001)	(0.000)	(0.807)	(0.731)	(0.591)	(0.915)	(0.042)	(0.784)	(0.995)	(0.008)	(0.409)	(0.766)	(0.511)
Interaction	0.266***	0.326***	-0.002	0.001	0.001	-0.001	-0.016	-0.005	-0.00003	0.010	-0.074	0.026	
	(0.000)	(0.000)	(0.808)	(0.755)	(0.727)	(0.915)	(0.173)	(0.784)	(0.995)	(0.193)	(0.420)	(0.767)	

Note: Table reports the wage differential and decomposition of the differential across the formally/informally employed (see first-stage equations in Table 3.9) showing the effects of characteristics, return and the interaction of characteristics and returns that explain the differential in wage across the formally/informally employed. Accounting for key variables as in Table 3.9. p-values in parentheses: * p<0.05; ** p<0.01; *** p<0.001. With clustered robust standard error, at the district level.

The Wage-Employed in urban Kenya

Evidence from Table 3.9 suggests the OLS estimate of returns to an additional year of schooling (continuous) for the public sector is 14.3% and that of the private-sector wage employed is 10.7%. However, considering returns across credential categories, evidence suggests only tertiary schooling has statistically significant returns (at the 5% level) in public service wage employment. Hence the returns to schooling for those working in the public sector with other credential categories is not different from nil. However, besides those with primary schooling, all other credential categories have positive (and statistically significant) wage effects. Interestingly, it is more beneficial to have post-primary schooling relative to secondary (and some post-secondary) schooling for private-sector wage employment in urban Kenya.

Table 3.12 shows the decomposition of the fitted model indicating a 91% wage differential for the public sector and private sector wage employed. Suggesting the model explains a 91% wage rise for the public sector over the private sector wage employed. About 46% of the wage differential is explained by differences in characteristics (of which differences in schooling make up over 95% of the differences in characteristics). Hence, differences in schooling between the public sector and private sector wage-employed explain about 44% of the wage differential between the two groups. Almost 54% (0.490/0.909) of the wage differential is attributable to unexplained potential discrimination in the reward for the public service over the private sector wage employed in urban Kenya. However, evidence suggests that the potential discrimination is partly explained by the high returns to the openness of the private sector over the public sector wage employed. To obtain more consistent estimates of the returns to an additional year of schooling, Table 3.14 suggests that the private sector wage-employed have a statistically significant return of about 18% in hourly earnings, and public sector wage-employed have a return to an additional year of schooling that is not different from nil, hence statistically insignificant. I now turn to the self-employed.

Table 3-12 Decomposition of Wage Differential, by Wage-Employment

		Differential											
		Prediction 1 (Private-Sector Wage-Employed)		0.555***									
				(0.000)									
		Prediction 2 (Public-Sector Wage-Employed)		1.464***									
				(0.000)									
		Difference		-0.909***									
				(0.000)									
	Total	YOS	Tenure	Age	BMI	Extra	Cons	Open	Sta	Agree	A_YOS	A_YOS_2	
Endowment	-0.483***	-0.460***	-0.009	-0.007	-0.028	0.004	0.001	-0.014	0.010	-0.008	0.137	-0.111	
	(0.000)	(0.000)	(0.741)	(0.723)	(0.172)	(0.500)	(0.934)	(0.391)	(0.311)	(0.438)	(0.258)	(0.308)	
Return	-0.490***	-0.501	0.081	0.312	-0.671	0.325	-0.028	0.348	0.606*	-0.430	1.431	-0.330	-1.510
	(0.000)	(0.299)	(0.427)	(0.422)	(0.137)	(0.294)	(0.957)	(0.354)	(0.041)	(0.173)	(0.661)	(0.829)	(0.404)
Interaction	0.064	0.115	-0.026	-0.020	0.028	-0.006	0.001	-0.015	-0.013	0.011	-0.029	0.012	
	(0.566)	(0.301)	(0.441)	(0.445)	(0.173)	(0.453)	(0.957)	(0.374)	(0.318)	(0.372)	(0.676)	(0.832)	

Note: Table reports the wage differential and decomposition of the differential across the Wage-Employed (see first-stage equations in Table 3.9) showing the effects of characteristics, return, and the interaction of characteristics and returns that explain the differential in wage across the Wage-Employed. Accounting for key variables as in Table 3.9. p-values in parentheses: * p<0.05; ** p<0.01; *** p<0.001. With clustered robust standard error, at the district level.

The Self-Employed in Urban Kenya

Across the self-employed, evidence from Table 3.9 suggests, the OLS estimate of the return to an additional year of schooling is 5.83% and 8.73% for the lone-employed and those in entrepreneurship respectively. Like the public-sector wage employed, the entrepreneurs only have statistically significant returns to their tertiary schooling. Whereas for the self-employed, in addition to the substantial returns to their tertiary schooling, they also have statistically significant returns to their non-degree technical schooling (isced4B), relative to those in entrepreneurship. Overall, the lone employed with primary, secondary, and some post-secondary educational attainments have no statistically significant returns to their schooling. Besides tertiary schooling (isced56) and some advanced technical non-tertiary schooling (isced4B), no other credential categories of education have statistically significant returns in self-employment. Most Kenyans in self-employment do not have tertiary (isced56) schooling and advanced technical non-tertiary (isced4B) schooling.

I now examine the wage gap between the lone employed and those in entrepreneurship. Evidence from Table 3.13 suggests a substantial wage differential of a 93% rise in hourly earnings for those in entrepreneurship relative to those in lone employment based on the fitted model. Further evidence from the decomposition analysis suggests the wage differential is mainly (about 65%) attributable to potential discrimination, suggesting this is partly due to the reward or returns to the extraversion of the entrepreneurs relative to the extraversion of the lone employed. The wage gap is also explained by differences in characteristics particularly, differences in schooling account for 75% of the effects of differences in characteristics on the wage gap. Turning to the evidence from the Two-Stage Least Square (2SLS-IV) approach, findings show a statistically significant return to an additional year of schooling for the entrepreneurs relative to the lone-employed in urban Kenya.

Table 3-13 Decomposition of Wage Differential, by Self-Employment

		Differential											
		Prediction1 (Entrepreneurship)		1.251***									
				(0.000)									
		Prediction 2 (Lone Employed)		0.321***									
				(0.000)									
		Difference		0.930***									
				(0.000)									
	Total	YOS	Tenure	Age	BMI	Extra	Cons	Open_av	Sta_av	Agree_av	Avg_Yos	Avg_Yos_2	
Endowment	0.183**	0.138**	0.007	-0.001	0.010	0.003	0.006	0.012	0.006	-0.001	-0.234	0.242	
	(0.004)	(0.002)	(0.497)	(0.729)	(0.430)	(0.730)	(0.592)	(0.295)	(0.445)	(0.839)	(0.129)	(0.125)	
Return	0.607***	0.279	0.160	0.565	0.451	0.829*	-0.432	-0.275	-0.752	0.240	-0.117	-0.035	-0.379
	(0.000)	(0.382)	(0.094)	(0.099)	(0.368)	(0.023)	(0.626)	(0.738)	(0.157)	(0.636)	(0.985)	(0.990)	(0.918)
Interaction	0.141	0.069	0.0302	0.010	0.021	0.045	-0.003	-0.008	-0.013	0.001	-0.009	-0.005	
	(0.243)	(0.395)	(0.216)	(0.582)	(0.418)	(0.074)	(0.715)	(0.746)	(0.418)	(0.843)	(0.985)	(0.990)	

Note: Table reports the wage differential and decomposition of the differential across the Self-Employed (see first-stage equations in Table 3.9) showing the effects of characteristics, return and the interaction of characteristics and returns that explain the differential in wage across the Self-Employed. Accounting for key variables as in Table 3.9. p-values in parentheses: * p<0.05; ** p<0.01; *** p<0.001. With clustered robust standard error, at the district level.

Table 3-14 2SLS-IV Estimates of Returns to Education and Skills

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	full	Female	Male	Informal	Formal	Public	private	Entr	Lone
	Panel A								
years_educ_act	0.063	-0.427	0.182*	0.072	0.327	0.115	0.201*	0.296*	-0.015
	(0.488)	(0.172)	(0.019)	(0.284)	(0.207)	(0.446)	(0.015)	(0.021)	(0.853)
avg_yos_	-0.128	0.331	-0.253*	-0.140	-0.452	-0.376	-0.232*	-0.394	0.139
	(0.309)	(0.310)	(0.039)	(0.190)	(0.066)	(0.172)	(0.033)	(0.683)	(0.329)
_cons	1.277	1.781	1.293	1.117	1.646	3.826	0.737	1.913	-0.836
	(0.075)	(0.469)	(0.130)	(0.147)	(0.447)	(0.301)	(0.307)	(0.849)	(0.488)
N	1771	753	1018	1272	498	138	923	121	578
R-sq	0.082	.	0.038	0.037	.	.	0.087	.	.
adj. R-sq	0.081	.	0.036	0.035	.	.	0.085	.	.
	Panel B								
apvlit_c	0.001	-0.006	0.005	0.004	-0.002	0.005	0.004	0.003	-0.001
	(0.665)	(0.188)	(0.089)	(0.259)	(0.747)	(0.613)	(0.194)	(0.825)	(0.712)
avg_skill_	-0.006	0.013	-0.012	-0.026	-0.007	0.024	0.001	0.110	0.004
	(0.780)	(0.451)	(0.544)	(0.594)	(0.679)	(0.566)	(0.951)	(0.141)	(0.842)
_cons	1.404	-0.780	1.843	4.231	2.935	-4.131	-0.296	-18.81	-0.151
	(0.669)	(0.800)	(0.571)	(0.598)	(0.305)	(0.536)	(0.898)	(0.122)	(0.967)
N	1780	759	1021	1277	502	140	928	121	580
R-sq	0.004	.	.	.	.	.	0.091	.	.
adj. R-sq	0.003	.	.	.	.	.	0.089	.	.

	Panel C								
years_educ_act	0.169	-0.293	0.256**	0.111	0.260	0.180	0.251**	0.218	0.008
	(0.072)	(0.231)	(0.005)	(0.120)	(0.269)	(0.148)	(0.006)	(0.055)	(0.932)
avg_yos_	-0.153	0.333	-0.246	-0.090	-0.343	-0.256	-0.226*	0.127	0.184
	(0.215)	(0.209)	(0.059)	(0.368)	(0.124)	(0.214)	(0.044)	(0.840)	(0.162)
Tenure	0.002	0.002	0.001	0.000004	0.005	0.004	0.0014	0.0113	-0.002
	(0.173)	(0.522)	(0.622)	(0.997)	(0.069)	(0.127)	(0.307)	(0.080)	(0.319)
tenure_sq	0.000004	0.000004	0.00001	0.00001*	-0.00002	-0.00001	0.000003	-0.00003	0.00002
	(0.355)	(0.796)	(0.144)	(0.015)	(0.171)	(0.303)	(0.506)	(0.215)	(0.069)
_cons	0.236	0.160	0.252	0.143	1.189	1.414	0.018	-3.286	-1.492
	(0.730)	(0.938)	(0.773)	(0.848)	(0.556)	(0.576)	(0.982)	(0.616)	(0.197)
N	1771	753	1018	1272	498	138	923	121	578
R-sq	0.142	.	.	0.087	0.096	0.215	0.002	0.098	0.026
adj. R-sq	0.140	.	.	0.084	0.088	0.191	-0.002	0.067	0.019
	Panel D								
apvlit_c	0.004	-0.004	0.008*	0.003	-0.001	0.006	0.005	0.003	-0.001
	(0.187)	(0.322)	(0.013)	(0.292)	(0.926)	(0.675)	(0.089)	(0.819)	(0.743)
avg_skill_	0.009	0.028	-0.003	0.013	0.0003	0.047	0.008	0.097	0.004
	(0.661)	(0.137)	(0.898)	(0.792)	(0.989)	(0.491)	(0.634)	(0.153)	(0.856)
Tenure	0.001	0.004*	-0.001	-0.0003	0.003	0.004	0.001	-0.009	-0.003
	(0.467)	(0.043)	(0.548)	(0.907)	(0.164)	(0.501)	(0.581)	(0.627)	(0.226)
tenure_sq	0.00001	-0.000004	0.00001*	0.00001	-0.00001	-0.00001	0.00001	0.00004	0.00002*
	(0.135)	(0.663)	(0.011)	(0.088)	(0.358)	(0.793)	(0.307)	(0.544)	(0.041)
_cons	-1.725	-3.748	-0.296	-2.484	1.183	-8.511	-1.789	-16.34	-0.075
	(0.620)	(0.249)	(0.933)	(0.771)	(0.703)	(0.427)	(0.511)	(0.120)	(0.985)
N	1780	759	1021	1277	502	140	928	121	580
R-sq	0.123	.	0.015	0.067	.	.	0.079	.	.
adj. R-sq	0.121	.	0.011	0.064	.	.	0.075	.	.

Note: Table reports outputs of variants of Equations 3.5 for the 2SLS-IV instrumenting schooling (years_educ_act) and skill (apvlit_c, not standardised); and district level schooling (avg_yos) and skill (avg_skill_). With the interaction of the reform dummy, p1985_ and quarters of birth (Q2, Q3, and Q4 with Q1 as a reference), as columns 1-9 with the outcome as log hourly earnings in USD. Columns (1) for the pool, (2) for females, (3) for males, (4) for the informal, (5) for the formal, (6) for the public-sector wage employed, (7) for the private-sector wage employed, (8) for entrepreneurs, (9) for the lone-employed. The predictors of interest are the continuous measure of schooling, years_educ_act and skill. Panels A and B are for schooling and skill respectively, with no controls. Panels C and D are for schooling and skill respectively, controlling for the number of months of experience in the current job (tenure) that also enters as a quadratic. See Tables 3.15 and 3.16 for the first-stage equations for schooling and skill respectively. p-values in parentheses: * p<0.05; ** p<0.01; ***p<0.001.

Table 3-15 First Stage of the 2SLS-IV (1)

	(1) Pool	(2) female	(3) Male	(4) inform	(5) Formal	(6) Public	(7) Private	(8) Entr	(9) Lone
	Panel A: years_educ_act								
0.p1985_#1.qob2	-0.042	-0.101	0.165	0.318	-0.619	0.150	-0.657	3.143	0.113
	(0.938)	(0.904)	(0.817)	(0.648)	(0.495)	(0.931)	(0.472)	(0.105)	(0.911)
1.p1985_#0.qob2	1.734***	1.421*	2.312***	2.502***	0.355	0.998	1.601*	1.530	2.041*
	(0.000)	(0.049)	(0.000)	(0.000)	(0.659)	(0.529)	(0.041)	(0.398)	(0.034)
1.p1985_#1.qob2	1.925***	1.376	2.759***	2.642***	0.747	1.892	1.924*	0.705	1.891
	(0.000)	(0.090)	(0.000)	(0.001)	(0.447)	(0.340)	(0.039)	(0.777)	(0.108)
0.p1985_#1.qob3	-0.085	0.813	-0.557	0.023	0.472	2.000	-0.898	4.800*	-0.994
	(0.888)	(0.416)	(0.463)	(0.975)	(0.662)	(0.274)	(0.362)	(0.027)	(0.372)
1.p1985_#0.qob3	-0.586**	-0.573*	-0.548	-0.714*	0.148	-0.442	-0.670	0.006	-0.330
	(0.004)	(0.034)	(0.070)	(0.022)	(0.703)	(0.579)	(0.063)	(0.996)	(0.490)
0.p1985_#1.qob4	-0.040	0.043	0.049	-0.566	0.381	-0.536	-1.378	3.000	4.40e-14
	(0.950)	(0.964)	(0.955)	(0.528)	(0.710)	(0.741)	(0.237)	(0.246)	(1.000)
1.p1985_#0.qob4	-0.267	0.092	-0.708*	-0.623	-0.536	-1.119	-0.470	-0.336	-0.121
	(0.203)	(0.737)	(0.026)	(0.058)	(0.205)	(0.161)	(0.225)	(0.770)	(0.817)
_cons	10.02***	9.687***	10.18***	9.066***	13.62***	14.25***	10.78***	11.00***	8.857***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	3061	1605	1456	1387	516	141	954	133	629
R-sq	0.012	0.010	0.022	0.024	0.011	0.037	0.019	0.065	0.032
adj. R-sq	0.010	0.006	0.018	0.019	-0.002	-0.014	0.012	0.013	0.021
Panel B: avg_yos_									
0.p1985_#1.qob2	0.151	0.403	-0.109	0.519*	-0.403	-0.097	-0.286	-0.021	0.964**
	(0.458)	(0.200)	(0.691)	(0.042)	(0.385)	(0.918)	(0.420)	(0.972)	(0.008)
1.p1985_#0.qob2	1.274***	1.262***	1.267***	1.733***	1.165**	0.871	1.802***	0.287	1.488***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.005)	(0.314)	(0.000)	(0.602)	(0.000)
1.p1985_#1.qob2	1.425***	1.351***	1.490***	1.856***	1.400**	0.878	1.978***	0.137	1.706***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.005)	(0.417)	(0.000)	(0.856)	(0.000)
0.p1985_#1.qob3	0.036	0.018	0.042	0.026	0.464	0.181	-0.191	0.470	0.093
	(0.874)	(0.963)	(0.885)	(0.926)	(0.401)	(0.856)	(0.618)	(0.471)	(0.817)
1.p1985_#0.qob3	-0.140	-0.282**	0.009	-0.140	-0.179	0.126	-0.359*	0.0210	0.059
	(0.069)	(0.006)	(0.942)	(0.222)	(0.365)	(0.771)	(0.010)	(0.950)	(0.734)
0.p1985_#1.qob4	1.334***	1.104**	1.525***	1.539***	1.043*	1.012	1.521***	-0.301	1.663***
	(0.000)	(0.002)	(0.000)	(0.000)	(0.047)	(0.254)	(0.001)	(0.701)	(0.000)
1.p1985_#0.qob4	-0.295***	-0.205*	-0.406***	-0.527***	-0.515*	-0.381	-0.605***	-0.128	-0.397*
	(0.000)	(0.047)	(0.001)	(0.000)	(0.017)	(0.379)	(0.000)	(0.713)	(0.037)
_cons	9.524***	9.588***	9.493***	9.136***	10.11***	10.03***	9.452***	10.67***	9.021***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	3078	1615	1463	1392	520	143	959	133	631
R-sq	0.037	0.023	0.060	0.074	0.040	0.039	0.079	0.015	0.067
adj. R-sq	0.035	0.019	0.055	0.069	0.027	-0.010	0.073	-0.041	0.057

Note: Table reports first-stage equation for the 2SLS-IV instrumenting schooling (years_educ_act) and average schooling at the district level (avg_yos_) using the interaction of the reform dummy, p1985_ and quarters of birth (Q2, Q3 and Q4 with Q1 as a reference), as columns 1-9. Columns (1) for the pool, (2) for female, (3) for male, (4) for the informal, (5) for the formal, (6) for the public-sector wage employed, (7) for the private-sector wage employed, (8) for entrepreneurs, (9) for the lone-employed. Panels A and B are for schooling and average schooling respectively. See Table 3.14 for the second-stage equations. p-values in parentheses: * p<0.05; ** p<0.01; ***p<0.001.

Table 3-16 First Stage of 2SLS-IV (2)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	pool	female	male	informal	formal	Pub	private	Entr	Lone
	Panel A: apvlit_c								
0.p1985_#1.qob2	-4.184	-20.44	8.916	-12.48	2.473	3.205	-3.286	57.98	-30.72
	(0.720)	(0.281)	(0.545)	(0.394)	(0.912)	(0.947)	(0.865)	(0.130)	(0.141)
1.p1985_#0.qob2	11.43	-11.10	34.94**	12.05	-5.010	2.903	-7.270	29.88	26.60
	(0.247)	(0.500)	(0.005)	(0.362)	(0.800)	(0.947)	(0.660)	(0.404)	(0.183)
1.p1985_#1.qob2	13.41	-13.37	41.66**	11.10	-8.265	12.82	-14.46	-19.65	34.20
	(0.251)	(0.470)	(0.007)	(0.491)	(0.732)	(0.815)	(0.463)	(0.690)	(0.161)
0.p1985_#1.qob3	-22.28	-14.42	-26.88	-29.85	6.875	17.29	-28.30	2.023	-29.06
	(0.091)	(0.528)	(0.088)	(0.059)	(0.796)	(0.732)	(0.175)	(0.962)	(0.209)
1.p1985_#0.qob3	-9.057*	-7.013	-10.47	-17.02**	18.59	15.53	0.825	-0.341	-26.64**
	(0.041)	(0.255)	(0.095)	(0.010)	(0.051)	(0.479)	(0.914)	(0.987)	(0.007)
0.p1985_#1.qob4	-38.17**	-53.33*	-25.62	-66.58***	5.237	0.687	-46.78	3.102	-68.34*
	(0.006)	(0.014)	(0.154)	(0.000)	(0.835)	(0.988)	(0.058)	(0.951)	(0.010)
1.p1985_#0.qob4	2.770	14.73*	-12.07	1.039	12.72	-1.595	18.86*	-12.72	-6.885
	(0.544)	(0.018)	(0.068)	(0.880)	(0.220)	(0.942)	(0.021)	(0.575)	(0.525)
_cons	180.1***	184.4***	178.0***	177.3***	203.8***	210.6***	183.7***	186.7***	175.2***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	3078	1615	1463	1392	520	143	959	133	631
R-sq	0.011	0.014	0.020	0.025	0.014	0.022	0.018	0.091	0.041
adj. R-sq	0.009	0.010	0.015	0.020	0.001	-0.029	0.011	0.040	0.031
Panel B: avg_skill									
0.p1985_#1.qob2	-0.937	0.776	-2.541	-0.114	-0.834	9.161	3.010	2.736	-4.594
	(0.768)	(0.880)	(0.535)	(0.977)	(0.909)	(0.574)	(0.584)	(0.800)	(0.386)
1.p1985_#0.qob2	1.417	-2.247	5.788	2.610	9.972	9.214	8.762	-3.303	-0.103
	(0.599)	(0.613)	(0.095)	(0.456)	(0.121)	(0.536)	(0.063)	(0.745)	(0.984)
1.p1985_#1.qob2	2.459	-3.048	8.962*	2.449	13.63	10.84	8.313	-4.928	2.853
	(0.440)	(0.543)	(0.036)	(0.566)	(0.083)	(0.560)	(0.139)	(0.724)	(0.646)
0.p1985_#1.qob3	-3.113	-5.842	-1.617	-3.176	-2.891	1.008	-3.538	7.237	-5.881
	(0.387)	(0.344)	(0.712)	(0.448)	(0.738)	(0.953)	(0.552)	(0.547)	(0.317)
1.p1985_#0.qob3	-1.047	-0.780	-1.269	-0.929	-1.854	0.0963	-3.001	-0.792	0.582
	(0.386)	(0.640)	(0.466)	(0.593)	(0.549)	(0.990)	(0.166)	(0.897)	(0.817)
0.p1985_#1.qob4	1.918	-1.735	5.404	-0.190	9.153	2.618	8.455	-7.241	0.076
	(0.612)	(0.768)	(0.279)	(0.970)	(0.265)	(0.864)	(0.228)	(0.616)	(0.991)
1.p1985_#0.qob4	0.567	3.564*	-3.141	-0.123	-4.125	-4.285	-0.507	0.407	-0.974
	(0.649)	(0.035)	(0.087)	(0.946)	(0.222)	(0.565)	(0.827)	(0.950)	(0.723)
_cons	174.1***	174.1***	174.0***	172.6***	175.7***	174.0***	171.4***	180.7***	173.6***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	3078	1615	1463	1392	520	143	959	133	631
R-sq	0.002	0.006	0.008	0.003	0.014	0.010	0.011	0.018	0.010
adj. R-sq	-0.000	0.001	0.003	-0.002	0.000	-0.042	0.003	-0.038	-0.001

Note: Table reports first-stage equation for the 2SLS-IV instrumenting skill (apvlit_c, not standardised) and average skill at the district level (avg_skill_, not standardised) using the interaction of the reform dummy, p1985_ and quarters of birth (Q2, Q3, and Q4 with Q1 as a reference), as columns 1-9. Columns (1) for the pool, (2) for females, (3) for males, (4) for the informal, (5) for the formal, (6) for the public-sector wage employed, (7) for the private-sector wage employed, (8) for entrepreneurs, (9) for the lone-employed. Panels A and B are for schooling and average skills, respectively. See Table 3.14 for the second-stage equations. p-values in parentheses: * p<0.05; ** p<0.01; ***p<0.001.

3.3.2 What is the (Private) Return to Skill in Urban Kenya?

In this subsection, the effects of skills (cognitive) are examined. Particularly, the objective is to assess how skill differs from schooling as a measure of human capital and to re-examine the arguments of Pritchett (2001) and Hanushek et al., (2015) that argue schooling may be an inappropriate measure of human capital as schooling, at best, result in minimal skills (due to quality issues) for productivity in developing contexts.

Return to Skill in Urban Kenya

Table 3.17 presents OLS (baseline) estimates of returns to skill in urban Kenya. Overall, evidence suggests consistency (in the trend of similar variables accounted for) using quantitative measures of schooling (years of education and credential categories) as measures of human capital. Models have a lower coefficient of determination. Hence, in using skill instead of schooling, models have lower explanatory powers of wage. Therefore, education causes more variation in hourly wage relative to skills. The effects of work experience and age on hourly wage are almost identical, regardless of the measures of human capital used. Relative to the use of schooling, the effect of gender enters more strongly, in that, they remain negative and statistically significant regardless of the use of continuous or categorical measures of skills. The overall wage effect of BMI is consistent with both measures of human capital. However, in accounting for non-cognitive skills, the use of schooling and skill are substantially different. The wage effects of Extraversion and Openness (which are the traits with statistically significant wage effects using both measures of human capital) are different in their mean coefficients and in their statistical significance using measures of schooling and skill.

Comparing columns (11) – (14) of Tables 3.8 and 3.17, the variability in the estimates of the wage effects of Extraversion and Openness suggests, the latter (Table 3.17) which accounts for cognitive skills and shows higher mean coefficients and more robust statistical significance is spurious. However, the former (Table 3.8) which accounts for schooling and shows more modest mean coefficients and statistical significance are consistent with the baseline wage effects of non-cognitive skills (see columns (1)-(7) of Table 3.23). Hence, evidence suggests accounting for cognitive and non-cognitive skills in a wage equation presents spurious or inconsistent outcomes, this is due to the correlation between cognitive and non-cognitive skills. Hence, to examine the wage effects of non-cognitive skills, I concentrate on the outputs of Table 3.8 which excludes measures of cognitive skills and to assess the wage effects of cognitive skills, I focus on outputs of Tables 3.17 and 3.18 which excludes non-cognitive skills in examining estimates of returns to skill. From columns (1)-(10) of Table 3.17, the evidence from the OLS estimate of returns to skill suggests a unit rise in the PV (Plausible Value) of an individual's cognitive skill explains an increase in hourly earnings of between (0.364 to 0.41) % using a continuous measure of skills. Turning to categorical skills, with the PV range of 0-175 (see Data Section) as a reference category, findings are consistent with the use of schooling as a measure of human capital, in that, higher levels of skill have higher returns. However, skills enter more strongly, in that magnitudes of mean coefficients across categories of skills are 'more progressive', meaning there is no case where lower category skills have higher returns relative to the subsequent skills level. Unlike the use of measures of schooling, here the returns to isced2 and isced34A categories of schooling show substantial inconsistency in their returns (wage effects), which further suggests material differences in skills between these groups (see Oledibe, 2023a (forthcoming). However, consistent with schooling as a measure of human capital, the return to cognitive skill is not different from nil, affirming the wage effect of human capital in urban Kenya is not statistically significant. I now examine heterogeneity in returns to skill and wage differentials across subsamples of interest.

Table 3-17 OLS (Baseline) Estimates of Returns to Cognitive Skill (Pool)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	ln_earnings_h_usd													
apvlit_c	0.004***		0.004***		0.004***		0.004***		0.004***		0.004***		0.004***	
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
apvlit_d_2		0.188*		0.167*		0.143		0.172*		0.145		0.179*		0.158*
		(0.024)		(0.029)		(0.074)		(0.023)		(0.066)		(0.010)		(0.031)
apvlit_d_3		0.679***		0.678***		0.615***		0.677***		0.616***		0.623***		0.572***
		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)
apvlit_d_4		1.111***		1.123***		1.011***		1.109***		1.004***		1.060***		0.965***
		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)
apvlit_h_5		2.002***		1.942***		1.951***		1.934***		1.950***		1.991***		2.006***
		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)
Tenure	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002**	0.002***	0.002***	0.002***	0.001**	0.002***	0.001**	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.003)	(0.000)	(0.003)	(0.001)
Age							0.006*	0.004	0.006*	0.005	0.007*	0.005	0.007*	0.006*
							(0.034)	(0.140)	(0.030)	(0.066)	(0.024)	(0.103)	(0.018)	(0.044)
Gender							-0.143*	-0.130*	-0.131*	-0.121*	-0.139*	-0.125*	-0.131*	-0.121*
							(0.011)	(0.023)	(0.017)	(0.033)	(0.015)	(0.030)	(0.018)	(0.034)
Bmi											0.015*	0.015*	0.016*	0.017*
											(0.020)	(0.026)	(0.012)	(0.013)
extraversion_av											0.085*	0.088*	0.077*	0.070*
											(0.032)	(0.018)	(0.040)	(0.048)
conscientiousness_avg											0.076	0.081	0.073	0.077
											(0.207)	(0.179)	(0.197)	(0.180)
openness_av											0.238***	0.235***	0.218***	0.213***
											(0.000)	(0.000)	(0.000)	(0.000)
stability_av											0.054	0.047	0.057	0.052
											(0.332)	(0.408)	(0.296)	(0.352)
agreeableness_av											-0.045	-0.056	-0.035	-0.043
											(0.292)	(0.202)	(0.395)	(0.302)
stratum_N			0.113	0.050			0.147	0.076			0.135	0.064		
			(0.163)	(0.511)			(0.078)	(0.321)			(0.120)	(0.432)		
stratum_L			-0.156*	-0.234**			-0.121	-0.207**			-0.149	-0.234**		
			(0.034)	(0.001)			(0.116)	(0.006)			(0.058)	(0.002)		
stratum_M			-0.007	-0.081			0.0224	-0.059			-0.003	-0.085		
			(0.928)	(0.296)			(0.776)	(0.448)			(0.968)	(0.263)		
avg_skill_					0.005***	0.006***			0.005***	0.005***			0.005***	0.005***
					(0.000)	(0.000)			(0.000)	(0.000)			(0.000)	(0.000)
_cons	-0.230**	0.143*	-0.085	0.367***	-1.021***	-0.698**	-0.218	0.286	-1.107***	-0.766**	-1.751***	-1.236***	-2.551***	-2.181***
	(0.004)	(0.019)	(0.409)	(0.000)	(0.000)	(0.004)	(0.176)	(0.055)	(0.000)	(0.003)	(0.000)	(0.001)	(0.000)	(0.000)
N	2234	2206	2234	2206	2234	2206	2234	2206	2234	2206	2124	2096	2124	2096
R-sq	0.104	0.115	0.113	0.127	0.116	0.128	0.120	0.131	0.122	0.133	0.150	0.161	0.150	0.160
adj. R-sq	0.104	0.113	0.111	0.123	0.115	0.126	0.116	0.127	0.119	0.129	0.145	0.154	0.145	0.154

Note: Table reports outputs of variants of Equations 3.3 and 3.4, as columns 1-14 with the outcome as log hourly earnings in USD. The predictors of interest are the continuous measure of skill, apvlit_c and categorical measures of skill, apvlit_d. Controls include strata-specific effects (based on the number of households across cities) and includes, Nairobi, strata_N, other large cities, strata_L with over 100 000HH, medium cities, strata_M with over 60 000HH but under 100 000HH, and other cities under, strata_S with under 60 000HH (reference category). Other control variables include the average skill, avg_skill, across districts, which also enters as a quadratic term. Other covariates include age, an indicator of female gender, tenure, or number of months in current employment and BMI. p-values in parentheses: * p<0.05; ** p<0.01; *** p<0.001. With clustered robust standard error, at the district level.

Heterogeneity Analysis Across Subsamples

OLS (Baseline) Returns

Table 3.18 presents the estimates of returns to continuous and categorical measures of cognitive skill across subsamples of interest, this also forms the first stage of the decomposition analysis that further examines wage differential and how differences in skill characteristics and return to skill impact wage differential across subsamples of the main analytical sample. Consistent with the use of schooling as a measure of human capital, the use of skill suggests returns to skill for females (of 0.418%) are higher than the return to skill for males (of 0.342%), interestingly, a striking consistency exists across categorical skills in genders where relative to males, females have higher return across all categories of skills. However, substantial inconsistency exists between each of the measures of human capital across the informally/formally employed. Here, evidence suggests the return to skill for the informal (of 0.301%) is greater, relative to those in formal employment (of 0.285%), this is consistent with findings using continuous and categorical skill measures. This suggests that the substantial informality in developing contexts adversely impacts the returns to human capital inhibiting growth. Interestingly, across the wage-employed, evidence suggests inconsistency in returns to skill relative to schooling. Here, the evidence suggests that the public service wage-employed have nil returns to their skill whereas the private sector wage employed have 0.323% returns to their cognitive skills, wage effects increase rising skill levels for the private sector wage employed. However, somewhat consistent with the categorical schooling measure, the evidence suggests that low-level skills have negative wage effects whereas, the highest skill levels have (substantially) high wage effects for the public service wage employed. Across the self-employed, evidence suggests no statistically significant wage effect of skill for those in entrepreneurship. However, for the lone employed, the evidence shows the most substantial returns to skills, with continuous schooling, the wage effect is a 0.372% rise in hourly wage which is the greatest across all employed. This substantial wage effect of skill for the lone employed (relative to other categories of the employed) is consistent with the categorical measure of skill. I now examine the wage differential, in the light of the differences in skill characteristics and the return effects (or potential discrimination) of skills, across the subsamples of interest in this study.

Table 3-18 OLS Return Estimates, by Genders and Employment Categories

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	Log Hourly Earnings in USD																	
	Pool		Gender (Female (1)/Male (0))				Informal (1)/Formal (0)				Wage (Public-Sector (1)/Private-Sector (0))				Self-Employed (Lone (1)/Entrepreneurship (0))			
	Continuous	Categorical	1		0		1		0		1		0		1		0	
apvlit_c	0.004***		0.004***		0.003***		0.003***		0.003***		0.002		0.003***		0.004***		0.003	
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.182)		(0.000)		(0.000)		(0.074)	
		0.157*		0.265*		0.069		0.246***		-0.303*		-0.544*		0.057		0.337**		0.068
		(0.032)		(0.035)		(0.379)		(0.000)		(0.014)		(0.020)		(0.477)		(0.002)		(0.801)
		0.626***		0.767***		0.513***		0.536***		0.243*		-0.037		0.516***		0.679***		0.423
		(0.000)		(0.000)		(0.000)		(0.000)		(0.023)		(0.883)		(0.000)		(0.000)		(0.170)
		1.039***		1.107***		0.946***		0.874***		0.522***		0.245		0.957***		1.087***		0.600
		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.383)		(0.000)		(0.000)		(0.124)
		2.002***		1.996*		1.930***		1.965**		1.481***		1.739***		1.508*		2.623***		0
		(0.000)		(0.028)		(0.000)		(0.006)		(0.001)		(0.000)		(0.012)		(0.000)		(.)
tenure	0.002**	0.002***	0.001	0.001	0.002**	0.002**	0.001*	0.001*	0.001	0.001	0.001	0.001	0.002*	0.002*	0.001	0.001	0.003	0.004
	(0.001)	(0.000)	(0.180)	(0.105)	(0.005)	(0.005)	(0.024)	(0.012)	(0.157)	(0.175)	(0.545)	(0.472)	(0.012)	(0.010)	(0.605)	(0.580)	(0.075)	(0.050)
age	0.006*	0.005	0.015**	0.012*	0.001	0.0002	0.005	0.004	0.007	0.007	0.001	-0.007	0.009**	0.009*	0.0001	0.0002	0.020	0.017
	(0.029)	(0.076)	(0.008)	(0.023)	(0.887)	(0.950)	(0.094)	(0.176)	(0.300)	(0.275)	(0.953)	(0.630)	(0.009)	(0.017)	(0.981)	(0.960)	(0.070)	(0.162)
bmi	0.015*	0.015*	0.016	0.015	0.016	0.018*	0.011	0.012	0.022**	0.023**	0.029	0.028	0.002	0.003	0.010	0.011	0.022	0.019
	(0.026)	(0.029)	(0.050)	(0.069)	(0.050)	(0.036)	(0.180)	(0.159)	(0.003)	(0.004)	(0.122)	(0.116)	(0.850)	(0.690)	(0.290)	(0.289)	(0.233)	(0.341)
avg_skill_	0.006***	0.006***	0.006**	0.007**	0.005***	0.005**	0.004**	0.004**	0.005*	0.005*	0.005	0.005	0.005**	0.005**	0.004	0.005*	0.004	0.003
	(0.000)	(0.000)	(0.008)	(0.002)	(0.000)	(0.001)	(0.004)	(0.002)	(0.018)	(0.019)	(0.175)	(0.157)	(0.004)	(0.002)	(0.062)	(0.049)	(0.545)	(0.610)
_cons	-1.602***	-1.243***	-2.179***	-1.874***	-1.258**	-0.870*	-1.187***	-0.974**	-0.953	-0.507	-0.572	0.212	-1.079**	-0.762*	-1.240**	-1.051*	-1.254	-0.672
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.028)	(0.001)	(0.005)	(0.055)	(0.324)	(0.517)	(0.805)	(0.002)	(0.023)	(0.006)	(0.025)	(0.297)	(0.614)
N	2107	2107	894	894	1213	1213	1546	1546	560	560	156	156	1076	1076	716	716	146	146
R-sq	0.127	0.139	0.138	0.147	0.119	0.135	0.080	0.083	0.097	0.139	0.093	0.184	0.119	0.143	0.097	0.104	0.153	0.149
adj. R-sq	0.125	0.135	0.133	0.138	0.114	0.128	0.077	0.078	0.087	0.125	0.056	0.133	0.115	0.136	0.089	0.092	0.117	0.099

Note: Table reports outputs of a variant of Equation 3.4, as columns 1-18 for subsamples of interest (Gender (3-6); Informality of Employment (7-10); Wage-Employment (11-14); Self-Employment (15-18)). The outcome is the log hourly earnings in USD. The predictors of interest are the continuous measure of skill, apvlit_c (not standardised), and categorical measures of skill, apvlit_d. Control variables include the average skill, avg_skill, across districts. Other covariates include age, an indicator of female, tenure, or number of months in current employment, and BMI. p-values in parentheses: * p<0.05; ** p<0.01; *** p<0.001. With clustered robust standard error, at the district level.

Explaining the Wage Gap and Causal Effects of Skill (on Wage) as a Measure of Human Capital

Tables 3.19-3.22 present the decomposition of skill differential across subsamples of interest. To start with, Table 3.19 suggests the fitted models (see Table 3.18) explain a wage differential of 22.9% rise in wages for males (relative to females). When schooling and non-cognitive skills are accounted for as earlier discussed, the wage differential was 23.4%. This suggests that wage differential is consistent across genders regardless of the measure of human capital used. Particularly, for males, irrespective of the measure of human capital, the evidence suggests that the fitted models explain a 68.8% rise in wage, however, for females accounting for schooling and non-cognitive skills explain a wage rise of 45.4% but accounting for cognitive skills instead of schooling, and non-cognitive skills result in a 45.9% rise in hourly wage. Interestingly, the use of skill unravels new insights as evidence suggests (although less precisely estimated) that potential discrimination explains substantially, the wage differential across genders. However, differences in characteristics (for which differences in skill constitute about 72% of differences in characteristics) explain about 37% of wage differential. Hence, skill differences alone explain 26.3% of the wage differential across genders.

Furthermore, consistent with estimates of returns to schooling relative to females, the males have statistically significant returns to their skills. A unit rise in PV of cognitive skill is causal to up to a 0.9% rise in hourly wage for males whereas, for the females, no statistically significant effects are attributable to the skills in urban Kenya. Turning to heterogeneity in returns across the formally/informally employed, Table 3.20 suggests the fitted models (see Table 3.18) explain a wage differential of a rise of 86.8% in hourly wage for the formal (relative to the informal). When schooling and non-cognitive skills are accounted for as earlier discussed, the wage differential was 87%. This suggests the wage differential is consistent across the formally/informally employed regardless of the measure of human capital accounted for. Somewhat inconsistent with the use of schooling as a measure of human capital where evidence suggests differences in characteristics best explain the wage gap between the formally and informally employed, using skill suggests the wage gap is mainly attributable to potential discrimination as 77.4% of the wage gap is attributable to potential discrimination. However, differences in characteristics (for which differences in skill constitute about 76.3% of differences in characteristics) explain about 22.4% of the wage differential. Hence, only 17.1% of the wage differential across the formally/informally employed is explained by skill differences.

Moreover, consistent with estimates of returns to schooling where evidence suggests the (causal) returns to schooling for the formal is statistically insignificant relative to the informal, the evidence suggests, no statistically significant (causal) effects of skills for both the informal and formal. I now turn to the wage employed. Table 3.21 suggests that the fitted models (see Table 3.18) explain a wage differential of 91.1% fall in hourly wage for the private-sector wage-employed (relative to the public-sector wage-employed). When schooling and non-cognitive skills are accounted for as earlier discussed, the wage differential was 90.9%. This suggests that the wage differential is consistent across the wage-employed irrespective of the measure of human capital accounted for. Consistent with schooling as a measure of human capital, accounting for skills suggests that 77.6% and 16.3% of the wage gap between the private-sector and public-sector wage employed are attributable to potential discrimination and differences in characteristics respectively.

In addition, consistent with estimates of returns to schooling, relative to the public-sector wage-employed, the evidence suggests that only the private-sector wage-employed have statistically significant (causal) returns to their skill. Hence, a unit rise in PV of cognitive skill is (causal) to a 0.6% rise in hourly wage for the private sector wage employed however, no statistically significant effects are attributable to the skills of the public-sector wage-employed in urban Kenya. I now turn to the self-employed. Table 3.22 suggests that the fitted models (see Table 3.18) explain a wage differential of 92.7% rise in wages for entrepreneurs (relative to the lone employed). When schooling and non-cognitive skills are accounted for as earlier discussed, the wage differential was 93%. Suggesting wage differential is consistent across the self-employed, regardless of the measure of human capital used. Consistent with schooling as a measure of human capital, the use of skill suggests about 84% of the wage gap is attributable to potential discrimination, although less precisely estimated, differences in characteristics (mainly due to differences in skill) only explain about 14% of the wage gap between those in entrepreneurship and those in lone employment. Finally, inconsistent with estimates of returns to schooling, no statistically significant causal returns exist for entrepreneurs, in urban Kenya.

Table 3-19 Decomposition of Wage Differentials, by Genders

Differential							
		Prediction 1 (Male)			0.688***		
					(0.000)		
		Prediction 2 (Female)			0.459***		
					(0.000)		
		Difference			0.229***		
					(0.000)		
	Total	Apvlit_c	Tenure	Age	BMI	Avg_Skill	
Characteristic/Endowment	0.084***	0.060***	0.004	0.009	-0.007	0.019	
	(0.000)	(0.001)	(0.247)	(0.201)	(0.160)	(0.085)	
Return/Coefficients	0.156**	-0.164	-0.157	-0.341	0.086	-0.037	0.921
	(0.003)	(0.065)	(0.137)	(0.123)	(0.742)	(0.932)	(0.089)
Interaction/Joint Effects	-0.011	-0.011	0.004	-0.008	-0.0002	-0.001	
	(0.454)	(0.164)	(0.423)	(0.236)	(0.967)	(0.932)	

Note: Table reports the wage differential and decomposition of the differential across genders (see first-stage equations in Table 3.18) showing the effects of characteristics, return, and the interaction of characteristics and returns that explain the differential in wage across genders. Accounting for key variables is shown in Table 3.18. p-values in parentheses: * p<0.05; ** p<0.01; *** p<0.001. With clustered robust standard error, at the district level.

Table 3-20 Decomposition of Wage Differential, by Formal/Informal

Differential							
		Prediction 1 (Formal)			1.229***		
					(0.000)		
		Prediction 2 (Informal)			0.361***		
					(0.000)		
		Difference			0.868***		
					(0.000)		
	Total	Apvlit_c	Tenure	Age	BMI	Avg_Skill	
Characteristic/Endowment	0.194***	0.148***	0.010	0.002	0.001	0.032*	
	(0.000)	(0.000)	(0.100)	(0.511)	(0.556)	(0.022)	
Return/Coefficients	0.672***	-0.027	0.005	0.063	0.277	0.143	0.234
	(0.000)	(0.782)	(0.917)	(0.784)	(0.193)	(0.675)	(0.681)
Interaction/Joint Effects	0.002	-0.008	0.001	0.001	0.002	0.007	
	(0.940)	(0.782)	(0.917)	(0.798)	(0.559)	(0.677)	

Note: Table reports the wage differential and decomposition of the differential across the formally/informally employed (see first-stage equations in Table 3.18) showing the effects of characteristics, return and the interaction of characteristics and returns that explain the differential in wage across genders. Accounting for key variables as in Table 3.18. p-values in parentheses: * p<0.05; ** p<0.01; *** p<0.001. With clustered robust standard error, at the district level.

Table 3-21 Decomposition of Wage Differential, by Wage Employment

Differential							
		Prediction 1 (Private Sector)			0.559***		
					(0.000)		
		Prediction 2 (Public Sector)			1.470***		
					(0.000)		
		Difference			-0.911***		
					(0.000)		
	Total	Apvlit_c	Tenure	Age	BMI	Avg_Skill	
Characteristic/Endowment	-0.148*	-0.078	-0.017	-0.002	-0.031	-0.022	
	(0.037)	(0.173)	(0.541)	(0.952)	(0.144)	(0.266)	
Return/Coefficients	-0.707***	0.312	0.052	0.269	-0.696	0.009	-0.507
	(0.000)	(0.300)	(0.603)	(0.574)	(0.141)	(0.987)	(0.518)
Interaction/Joint Effects	-0.057	-0.060	-0.016	-0.017	0.029	-0.0002	
	(0.447)	(0.305)	(0.608)	(0.584)	(0.176)	(0.987)	

Note: The table reports the wage differential and decomposition of the differential across the wage-employed (see first-stage equations in Table 3.18) showing the effects of characteristics, return, and the interaction of characteristics and returns that explain the differential in wage across genders. Accounting for key variables is shown in Table 3.18. p-values in parentheses: * p<0.05; ** p<0.01; *** p<0.001. With clustered robust standard error, at the district level.

Table 3-22 Decomposition of Wage Differential, by Self-Employment

Differential							
		Prediction 1 (Entrepreneurship)			1.246***		
					(0.000)		
		Prediction 2 (Lone Employed)			0.319***		
					(0.000)		
		Difference			0.927***		
					(0.000)		
	Total	Apvlit_c	Tenure	Age	BMI	Avg_Skill	
Characteristic/Endowment	0.129**	0.090*	0.005	0.0001	0.011	0.024	
	(0.004)	(0.012)	(0.616)	(0.981)	(0.357)	(0.160)	
Return/Coefficients	0.777***	-0.167	0.151	0.665	0.304	-0.093	-0.014
	(0.000)	(0.521)	(0.154)	(0.051)	(0.532)	(0.931)	(0.991)
Interaction/Joint Effects	0.0215	-0.025	0.0282	0.010	0.014	-0.003	
	(0.708)	(0.532)	(0.260)	(0.599)	(0.554)	(0.931)	

Note: The table reports the wage differential and decomposition of the differential across the self-employed (see first-stage equations in Table 3.18) showing the effects of characteristics, return, and the interaction of characteristics and returns that explain the differential in wage across genders. Accounting for key variables is shown in Table 3.18. p-values in parentheses: * $p<0.05$; ** $p<0.01$; *** $p<0.001$. With clustered robust standard error, at the district level.

Table 3-23 Mediation Analysis of Schooling and Skills

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	ln_earnings_h_usd															
extra_av	0.054	0.071	0.055	0.071	0.059	0.034	0.050									
	(0.220)	(0.122)	(0.161)	(0.118)	(0.121)	(0.430)	(0.203)									
con_avg	0.126*	0.128	0.121*	0.123	0.115*	0.118*	0.115*									
	(0.046)	(0.051)	(0.041)	(0.053)	(0.046)	(0.047)	(0.046)									
open_av	0.359***	0.366***	0.307***	0.358***	0.303***	0.308***	0.298***									
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)									
sta_av	0.076	0.075	0.079	0.060	0.065	0.059	0.063									
	(0.137)	(0.137)	(0.091)	(0.253)	(0.180)	(0.231)	(0.190)									
agree_av	-0.046	-0.062	-0.046	-0.060	-0.048	-0.051	-0.050									
	(0.349)	(0.220)	(0.327)	(0.231)	(0.300)	(0.281)	(0.277)									
apvlit_c								0.004***		0.004***						0.001**
								(0.000)		(0.000)						(0.010)
apvlit_d_2									0.212*		0.168					
									(0.018)		(0.064)					
apvlit_d_3									0.703***		0.629***					
									(0.000)		(0.000)					
apvlit_d_4									1.129***		1.008***					
									(0.000)		(0.000)					
apvlit_h_5									1.943***		1.947***					
									(0.000)		(0.000)					
years_educ												0.112***		0.107***		0.093***
												(0.000)		(0.000)		(0.000)
isced1													0.037		0.027	
													(0.629)		(0.712)	
isced2													0.438***		0.477***	
													(0.001)		(0.000)	
isced34A													0.470***		0.438***	
													(0.000)		(0.000)	
isced4B													1.110***		1.073***	
													(0.000)		(0.000)	
isced56													1.728***		1.688***	
													(0.000)		(0.000)	
Tenure	0.002***	0.002***	0.002***	0.002***	0.002***	0.002**	0.002***	0.002***	0.002***	0.002***	0.002***	0.0020***	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age				0.003	0.004	0.006	0.004	0.007*	0.005	0.007*	0.005	0.0076*	0.005	0.010**	0.006*	0.009**
				(0.482)	(0.260)	(0.101)	(0.215)	(0.045)	(0.178)	(0.040)	(0.092)	(0.024)	(0.152)	(0.003)	(0.042)	(0.007)
Gender				-0.156*	-0.139*	-0.155*	-0.143*	-0.144*	-0.139*	-0.133*	-0.128*	-0.112*	-0.099	-0.118*	-0.102	-0.101
				(0.012)	(0.019)	(0.011)	(0.016)	(0.017)	(0.028)	(0.023)	(0.039)	(0.041)	(0.069)	(0.032)	(0.058)	(0.057)

stratum_N		0.027		0.047				0.162*	0.082			-0.119	-0.052			
		(0.787)		(0.657)				(0.046)	(0.293)			(0.109)	(0.370)			
stratum_L		-0.245*		-0.225*				-0.119	-0.206*			-0.343***	-0.269***			
		(0.013)		(0.029)				(0.185)	(0.015)			(0.000)	(0.000)			
stratum_M		-0.020		-0.010				0.075	0.001			-0.190*	-0.146			
		(0.859)		(0.929)				(0.452)	(0.993)			(0.032)	(0.069)			
avg_skill_			0.008***		0.008***		0.006**			0.006***	0.006***					0.005**
			(0.000)		(0.000)		(0.003)			(0.000)	(0.000)					(0.003)
avg_yos_						-0.233*	-0.142							-0.344**	-0.170	-0.228*
						(0.033)	(0.215)							(0.001)	(0.079)	(0.036)
_avg_yos_sq_						0.016**	0.009							0.018***	0.009*	0.010
						(0.002)	(0.117)							(0.000)	(0.049)	(0.070)
_cons	-1.201***	-1.166***	-2.463***	-1.107**	-2.412***	-0.377	-1.521*	-0.430**	0.101	-1.298***	-0.970***	-0.728***	-0.042	0.723	0.597	-0.530
	(0.001)	(0.001)	(0.000)	(0.002)	(0.000)	(0.574)	(0.041)	(0.008)	(0.478)	(0.000)	(0.000)	(0.000)	(0.743)	(0.236)	(0.293)	(0.455)
N	1896	1896	1896	1896	1896	1896	1896	1902	1879	1902	1879	1892	1892	1892	1892	1892
R-sq	0.056	0.066	0.082	0.071	0.086	0.082	0.088	0.131	0.138	0.134	0.142	0.204	0.275	0.197	0.268	0.209
adj. R-sq	0.053	0.061	0.078	0.065	0.082	0.077	0.083	0.128	0.133	0.132	0.138	0.201	0.271	0.195	0.264	0.206

Note: Table reports outputs of Mediation Analysis involving Cognitive, Non-Cognitive Skills, and Schooling, for the latter—cognitive skill and schooling, I examine both the continuous and categorical measures. The outcome variable is the log of hourly earnings in USD. Additional variables include strata-specific (district-size) effects based of the stratified sampling (based on the number of households (HH) across cities). This categorises all districts into either of the following: Nairobi, strata_N; Other Large districts, strata_L with over 100 000HH; Medium districts, strata_M with over 60 000HH but under 100 000HH; and other districts, strata_S with under 60 000HH (reference category). Other variables include average schooling, avg_yos across districts which also enters as a quadratic term, and average skill (not standardised). Other covariates include tenure, age, and an indicator of female gender. p-values in parentheses: * $p<0.05$; ** $p<0.01$; *** $p<0.001$. With robust standard error, clustered at the district level.

3.4 Summary and Concluding Remarks

I now present a summary, limitations, implications, and the way forward (for future studies) of the findings of this study aimed at improving understanding of the growth effects and returns to education, skills, and health as measures of human capital in developing contexts.

To start with health and non-cognitive skills or personality traits. On the wage effect of heath as human capital. Although mean effects are minimal and less precisely estimated, evidence suggests individual health impacts individual earnings in urban Kenya as a unit rise in BMI has a wage effect of an average rise in hourly earnings of between (1.3 to 1.4) %, this is statistically significant at the 5% level. However, further evidence suggests only the male gender and those that are formally employed have wage effects on their health. On the wage effects of non-cognitive skills or personality traits, a unit rise in Openness to experience results in hourly wage rise of an average of (10.8-17) % in urban Kenya. This has strong effect, particularly, it is at least statistically significant at the 1% level. However, further evidence suggests only the males, the informally employed and those in private-sector wage employment have positive and statistically significant wage effects of their Openness, with effects strongest for males and private-sector wage employed with up to 24.5% and 21.5% rise in hourly wage for a unit rise in Openness for the male gender and private-sector wage employed respectively. Interestingly, only the female gender, the informally- and lone-employed have strong wage effects for their Conscientiousness.

Mean effects are greatest for the lone-employed but less precisely estimated as a unit rise in conscientiousness results in up to 24.3% rise in hourly wage for the lone-employed, statistically significant at the 5% level. However, for the females, mean coefficients are also considerably large but more precisely estimate, with a unit rise in conscientiousness resulting in up to 23% rise in hourly wage, statistically significant at the 1% level. Interestingly, whilst Agreeableness impacts the wage of the formally employed resulting in up to 15.2% wage rise for a unit increase in agreeableness, statistically significant at the 5% level. However, for the informally employed, Openness and Conscientiousness positively impact hourly wage, with a wage rise of 16.5% and 15.6% for a unit rise in Openness and Conscientiousness respectively. The wage effect of Openness is much stronger relative to Conscientiousness not only for the informally employed but in urban Kenya. Using the main analytical sample, Conscientiousness has no statistically significant effects on average hourly wage across respondents in urban Kenya. However, as earlier discussed Openness strongly impacts average hourly earnings in urban Kenya as evidence from the main analytical sample indicates an average wage rise of (10.8-17) % is attributable to Openness. This is statistically significant at the 1% level (at the very least). (These effects are deemed causal (effects) of personality traits or cognitive skill on wage).

In this subsection of the analysis, the main objective is to summarise returns to human capital and in doing so, examine consistency/inconsistency in the use of schooling and skill as measures of human capital that impact growth. The use of quantitative and qualitative schooling such as measures of educational attainment (e.g., years of schooling; and credentials); and cognitive skills respectively. Consistent with both measures of human capital, evidence suggests only the male gender and those in private-sector wage employment have a causal wage effect on an additional year of their schooling in urban Kenya. The evidence from 2SLS-IV indicates that, for the male gender, an additional year of schooling results in a 20% rise in hourly wage and a unit rise in the PV of their cognitive skill results in a 0.96% rise in hourly wage, both effects are statistically significant at the 1% level. For the private-sector wage employed, evidence suggests that an additional year of schooling results in a 19.9% rise in hourly wage, and a unit rise in the PV of cognitive skill results in a 0.652% rise in hourly wage, both effects are statistically significant at the 5% level. Interestingly, across genders, and the wage-employed, the female gender and the public-service wage employed have no significant wage effects of their schooling and skill on which causal inferences can be drawn. With an emphasis on the impact of schooling and skill characteristics using the Oaxaca-Blinder Decomposition, examining OLS (baseline) evidence of the wage differential across genders and the wage employed,

Across genders, about 23% wage differential exists for males over females. Of this wage gap, differences in human capital (schooling and skill characteristics) of the males over the females explain close to thirty-seven per cent (37.5%) of the wage gap. However, substantial evidence of potential discrimination exists across genders and one of the mechanisms for this is through ‘potential discrimination’ in the reward for (or returns to) personality traits across genders. Evidence suggests males have higher rewards for their Openness, relative to females whereas, for conscientiousness, the reverse is the case across genders. Across the wage-employed, evidence suggests a wage differential of 91% rise in hourly wage for the public sector over the private-sector wage employed. Interestingly, evidence suggests over fifty percent of the wage gap is attributable to potential discrimination and the reward or return to the stability of the private sector over the public sector wage employed partly explains this discrimination.

Whilst evidence suggests differences in schooling characteristics of the public sector over the private sector explain about 50% of the wage differential, further evidence suggests no statistically significant difference in skill characteristics across the wage employed. This inconsistency between schooling and skill aligns with the argument of Pritchett (2001), that argues cognitive skills may not be applied in socially productive active activities, particularly in public services in developing contexts. This has a substantial adverse economic growth effect and has less to do with skills from schooling, but a system of governance that discourages the use or application of skills needed for growth. Consistent with the argument of Pritchett (2001), findings reveal the public sector wage-employed have no positive causal effects of their human capital (schooling and skill) rather, a substantial (and consistent) causal wage effects of the human capital (both for skill and schooling) of the private-sector wage employed is evidenced in this analysis, although less precisely estimated, suggesting skills better capture true human capital. So far, this marks the first inconsistency between schooling and skills as measures of human capital. A more striking inconsistency between schooling and skill is seen in the wage effects of human capital across the formally/informally employed.

Evidence from the 2SLS-IV suggests skills have no statistically significant wage effect of the formal, however, an additional year of schooling has a ‘causal’ wage effect of a 33.6% rise in hourly wage, less precisely estimated with a 5% statistical significance. With a wage differential of about 87% rise in wages for the formal over the informal. Interestingly, differences in schooling and skills (human capital) characteristics substantially explain this wage gap. However, whilst potential discrimination through rewards for (or return to) schooling and Agreeableness trait for the formally employed over the informally employed, potential discrimination of the informally employed is not explained through reward for (returns to) skill possessed. However, the (favourable) reward for the Conscientiousness of the informally employed also accounts for potential discrimination between the formally/informally employed. Hence, the causal wage effect of schooling and nil effect of skill for the formally over the informally employed. This inconsistency in the ‘causal’ estimates of returns across the formal and informal strongly suggests that, whilst skill is a useful measure of human capital, schooling as a measure of human capital remains crucial especially in examining the growth effects (of schooling) and in obtaining consistent estimates of non-cognitive skill.

Evidence from the quasi-experimental analysis (2SLS-IV) suggests there are no statistically significant wage effects of human capital in urban Kenya. However, subsampling unravels useful insights as opposed to the main analytical sample, the males have substantial returns to their human capital, both across the measures of schooling and skill which suggests the robustness of this outcome. However, the mean effect of skill is weaker and less statistically significant, relative to education. With men having as high as a 35% rise in hourly earnings for an additional year of their schooling. Across the employed, those in entrepreneurship and the private-sector wage-employed show similar evidence as those of the male gender. Whilst most women are thought to be self-employed, it is reasonable to state that the returns to those in entrepreneurship and private-sector wage employment only reflect the (effects) of the returns to schooling and skill for the male gender. Effectively, no substantial evidence suggests females have statistically significant returns to their human capital in urban Kenya besides the OLS evidence that appears to overstate the returns for females. The evidence from the 2SLS-IV estimate is consistent with the evidence of potential discrimination and substantial inequality in schooling that adversely impact the returns to the schooling of the female gender.

Data limitations and hence, the inability to causally identify returns that fully account for credential categories, is a major limitation to this study. This is a subject that is to be considered in future works. The returns to an additional year of schooling should vary substantially across credential categories. Hence, the return to an additional year of schooling for a respondent with primary education should differ (substantially), compared with the return to an additional year of schooling for a respondent with tertiary education. However, in this analysis, the (causally) identified return estimates reported have not accounted for the credential categories of the respondents. Whilst the gender and employment categories that have causally identified return estimates are clearly stated, the causally identified return estimates stated apply to a subgroup of the gender and employment categories involved. The causally identified return estimates give a Local Average Treatment Effect (LATE), and this return estimate refers to those whose schooling and skill are impacted by the 1985 curriculum structural reform and the quarter of their birth. It is not clear who these individuals are among the gender and employment categories stated. Finally, this study has proxied human capital with schooling or cognitive skills, however, more recent studies now consider composites of these (see Araki (2020)) especially in the fields of Sociology, although using such composites presents further econometric complications, especially in obtaining causally identified return estimates of such. However, such would be a more appropriate measure that captures effective human capital in developing contexts. Future studies should consider these limitations and effectively improve knowledge in these lights.

4 Human Capital Externalities and Social Returns

4.1 Introduction

In this chapter, I examine human capital externalities—the effects of aggregate (district-level) schooling and skills on individual skills and earnings. Drawing from the previous chapter on the private returns to schooling, inspired by the study of Liu (2007), I provide estimates of Social Returns as an approximation of the external and private returns in urban Kenya.

4.1.1 The Problem and Objective of this Study

In 2017, the education system in Kenya was ascribed to a leading force in Africa, according to the World Bank and the World Economic Forum. As part of its strategy for economic growth, Kenya's immense investment and reforms in education, relative to most countries in sub-Sahara Africa are known to show the country's resolve to raise skills for human capital. Specifically, the launch of the competency-based curriculum in 2017 and the substantial education expenditure (as a % of its GDP)[92]. Such commitments and spending on education are far above the average of the sub-Sahara Africa region. However, what should justify this level of attention and (public) funding on education, amidst the relatively high private returns to education in sub-Sahara Africa?

With the high private returns[93] in Kenya, it is reasonable to argue that a strategy for more schooling over time will raise individual earnings and then, in aggregate, lead to economic prosperity. However, evidence from the growth accounting models suggests social returns[94] are inconsistent with the evidence from private returns. This suggests that the wage effects of an additional year of schooling for an individual (private return) do not necessarily translate to the average wage effects of the average schooling of the people (social returns). This presents a micro-macro conundrum that implies the growth impact of education varies. The study of Pritchett (2001) suggests that the developmental effects of education vary substantially across the countries of the world (also, see Temple, 1999). Whilst Temple

---

[92] https://data.worldbank.org/indicator/SE.XPD.TOTL.GD.ZS?end=2021&locations=ZG-KE&start=1971&view=chart. Kenya's spending as a % of GDP.

[93] Here, private return is defined as the percentage rise in hourly wage for an additional year of schooling of an individual (micro evidence). The study of Oledibe (2023b, forthcoming) finds this to be 18.2%; 20.1%; 29.6% for the male gender; the private sector waged employed; and the entrepreneurs.

[94] Here, social return is defined as average or aggregate wage increment to an additional year of average (or aggregate) schooling for an economy (macro evidence).

(1999) stressed the importance of technology on the growth effects of education, Pritchett (2001) puts forward three propositions[95] in an attempt to explain why the effects of schooling on economic growth fall short of expectations in developing contexts characterised by substantial decreases in (average) wage, amidst the rising (average) educational attainment (see Barro and Lee, (2013)).

The study of Pritchett (2001) asserts that the rising supply of educated capital that outstrips the demand for educated capital is a possible reason for a fall in returns in developing contexts. This suggests the need to account for measures of educational expansion or aggregate human capital in determining return estimates—the effects of these measures of aggregate human capital capture human capital externalities. It is well-founded in the literature (see Moretti (2006)) that this conundrum (or inconsistency) between micro and macro evidence of the returns to schooling is at least, in part, due to human capital externalities. The study of Moretti (2006) argues that the pecuniary and non-pecuniary growth impacts of such aggregate human capital (externalities) should give the basis for government funding of (tertiary) schooling, according to 'efficiency[96] advocates' (see Moretti, 2004; and 2006).

In this study, having considered private pecuniary and non-pecuniary returns in previous chapters (chapters 2 and 3)[97], inspired by the study of Moretti (2006) in developed contexts and the emphasis on skill in this study, I examine pecuniary (human capital) externalities—the effects of aggregate education (and skills) on individual earnings in urban Kenya. Most existing studies have only considered the pecuniary externalities of schooling. Moreover,

[95] Pritchett's (2001) propositions include the following: issues of quality of schooling where years of schooling produce minimal skills at best; fall in marginal returns to schooling attributable to the supply of educated labour that outstrips the demand for educated labour; the environment where skill from schooling may be privately remunerative but socially unproductive (see Pritchett, 2001).

[96] Whilst most countries of the world inclusive of the non-OECDs, now offer at least free primary education, it is important to note that some OECD countries go beyond covering most direct (like tuition and other ancillary) costs of schooling to providing for indirect (or opportunity) costs by offering stipends. In the US, public universities get funds and private universities are well funded by the Central Government (Moretti, 2006). Empirical research has attempted to justify government funding of post-primary education with two broad understandings – equity and efficiency (Moretti, 2006) – Justifying education funding with the understanding of the externalities of schooling is one of the ideas of the 'efficiency' advocates. Besides efficiency advocates, the 'equity' advocates are of the opinion of achieving a fair redistribution of wealth (which ordinarily, may not be achievable due to the market mechanism) to constrain the beneficiaries to acquire education in lieu of other goods, is socially desirable.

[97] This study is complementary to the previous chapters, please, see (Oledibe 2023a; 2023b) for chapters 2 and 3 respectively, where the effects of skill from schooling (a non-pecuniary private returns to schooling); and the wage effects of schooling and skills (a pecuniary private returns to human capital—schooling and skills) are analysed.

whilst the focus of this study is on the pecuniary externalities, in examining the effects of aggregate schooling on earnings (pecuniary externalities of schooling); relative to the effect of aggregate skills on earnings (pecuniary externalities of skills), it becomes useful to consider non-pecuniary externalities of schooling—such as the effects of aggregate education on the individual's cognitive skills. The consideration for non-pecuniary externalities of schooling may give useful insights into the relationship between the pecuniary externality of schooling and pecuniary externalities of skills.

Furthermore, inspired by the study of Liu (2007) in China, I draw on an approximation for pecuniary and non-pecuniary social returns to education and skills in urban Kenya. To estimate Social Returns[98], I do this by taking the sum of the external and private returns (see Liu 2007). These estimates of pecuniary and non-pecuniary human capital externalities and approximation of social returns for Kenya, make it possible to examine the claims (see propositions) of Pritchett (2001), providing empirical evidence in urban Kenya. Furthermore, these estimates for Kenya provide similar evidence of human capital externalities and social returns as those of Moretti (2006) for the United States of America; and Liu (2007) for China. Insights from these estimates provide a basis for an assessment an assessment of government intervention in the public provision of schooling (see Moretti (2006); and Liu (2007)).

4.1.1.1 The Research Questions

(1). What is the effect of aggregate education and aggregate skill on the wage of individuals? Hence, what is the pecuniary Social Return (to human capital), in urban Kenya?

(2). What is the effect of aggregate education on skills (of the individual) in urban Kenya? Hence, what is the non-pecuniary Social Return, in urban Kenya?

4.1.2 Antecedents—the Review of Related Literature.

The early study of Lucas (1988) attempts to model economic development using three models which include one that emphasises learning by doing, a model of specialised human

[98] Consequently, I examine the sum of the wage effects of an additional year of schooling for the individual; and the effects of an additional year of aggregate schooling in cities (as the external returns to schooling) on individual earnings—as an approximation of the pecuniary social returns to schooling. I do the same for skills to obtain an approximation of the pecuniary social returns to skills. With individual skill as an outcome, the sum of the effects of individual schooling and aggregate schooling as inputs give an approximation of the non-pecuniary social returns to schooling.

capital accumulation; a model of human capital accumulation through schooling; and lastly, a model of physical capital accumulation with technological change. Using these, Lucas (1988) asserts that in a city, having highly skilled/educated workers, generates benefits for other workers, suggesting that the average human capital in a city impacts the wage of the low-skilled/educated by raising their skills. However, the positive relationship between average human capital and average wage, although dependent on the quantity of physical capital, may very much be dependent on the quality of physical capital or technology (Moretti, 2006), suggesting the place of technology on the impact of average human capital on individuals (also, see, Temple, 1999). At the other extreme of this argument is the literature on skill-biased technical change, where a rise in average human capital or the increased supply of the educated/skilled labour raise the demand for the highly educated/skilled as investment in physical capital rises.

The complementarity of physical (technology or skill) and human capital is vital to making the distinctions between the effects of rise in human capital and returns that impact educated/skilled over the less educated/skilled. However, the study of Pritchett (2001) suggests, in developing contexts characterised with minimal technology (skill), with the supply (over demand) of educated labour, hence, a rise in average human capital adversely impact the returns to schooling. On the other hand, in technology-rich contexts (see Autor and Acemoglu (2011)), the study of Autor and Acemoglu (2011) is consistent with the literature of skill-biased technical change. They affirm technology result in lower demand for labour-intensive or the low-skilled/educated, hence, raising their supply over demand (Autor and Dorn, 2013; Autor et al., 2015), this adversely impacts the returns of the human capital of low-skilled/educated. With the basic understanding of market forces, it is rational to conclude that the returns to education should decrease with rising district-level educational attainment, hence, educational expansion should devalue human capital or have adverse wage effect. This is very much in line with Pritchett's (2001) argument in the non-OECDs with limited quantity and quality of physical capital (technology). Where he puts forward his thoughts (propositions) about rising supply of labour with limited demand resulting in fall of the returns to schooling.

It is crucial to take note of the difference in contexts. Whilst Pritchett (2001) referred to a non-OECD context characterised by lone-self-employment. Lone employment suggests low skills and raise supply of labour. On the other hand, the work of Acemoglu and Autor (2011) which was about a decade apart from that of Pritchett's (2001) is based in the OECDs, characterised by high growth entrepreneurship that raise demand for skills. Interestingly, the study of Berman et al. (1998) suggests that regardless of the contexts—developing or developed—Skill-Biased Technical Change (SBTC) means an increase in the premium for high skills prevail. I argue that whilst SBTC prevail around the world and results in premium for high skills, this may not apply in the same extent across developed and the developing countries as high unemployment even among the highly skilled in developing countries compel substantial emigration from developing countries. What is clear from the conclusions so far (Berman et al, 1998; Pritchett, 2001; Autor and Acemoglu, 2011) is the place of skills in the different labour markets and education systems along the OECDs and non-OECDs divide—by this, I mean the extent to which skill is rewarded in the respective labour markets and the extent to which schooling raise skills in both contexts differs substantially.

Skills appear to be largely misplaced in the non-OECDs—by this, I mean inefficiencies in education and a case where labour markets in developing contexts do not reward skills as due. This can substantially explain variability in returns based on skills across different contexts. However, skills (and technology) remain a crucial mechanism through which education impacts or result in growth. Hence, I argue that understanding how aggregate schooling (district-level schooling) impacts individual level skills; and in turn, how aggregate human capital (district-level schooling and skills) impacts individual earnings in urban Kenya will improve understanding of SBTC in urban Kenya and effectively examine the arguments of Pritchett (2001) in developing contexts.

In this study, the objective is to empirically examine the claims of Pritchett (2001) — that suggests rising district-level educational attainment minimises returns in the non-OECDs. This gives insights to the claims of Autor and Acemoglu (2011) in the OECDs that gave opposing views to those of Pritchett (2001) on the effects of aggregate human capital on returns of individuals. Ultimately, this will aid an assessment of the conclusion of Berman et al. (1998) across developing and developing contexts. Hence, beyond estimating and examining the pecuniary human capital externalities argued to be a useful basis for public funding of schooling (Moretti, 2006), I examine the effects of such external returns lending my voice to the literature on SBTC—emphasising the effects of these external returns on human capital, in the context of urban Kenya.

Contributions

There is a plethora of studies on private returns to education in sub-Sahara Africa, however, less attention has been devoted to the role of aggregate human capital on the skills and earnings of individuals as human capital externalities in sub-Sahara Africa. I contribute to the literature on human capital externalities in three areas, firstly, most studies on human capital externalities provide evidence in developed contexts (with substantial technological improvements) or other regions of the world besides sub-Sahara Africa. Using the World Bank's STEP data, I show evidence of human capital externalities for Kenya, in sub-Sahara Africa. Understanding human capital externalities for Kenya holds several benefits for policymaking on economic growth of Kenya across the districts (which make the provinces) of Kenya, where substantial variations in schooling and skills (see Table) exist across districts. Secondly, the only known (existing) study for human capital externalities for Kenya, is those of Kimenyi et al. (2006) who provide estimates of private returns and externalities of human capital taking both individual and average schooling as exogenous. In this study, exploiting the 1985 curriculum reform in Kenya that result in exogenous variation in schooling, I improve on internal validity front, taking both individual and aggregate schooling as endogenous, drawing causal inferences from estimates. Furthermore, beyond examining human capital externalities, this study examines non-pecuniary human capital externalities (effects of aggregate schooling on individual skill) using this as possible mechanisms that explain differences in the pecuniary externalities of schooling and skill in urban Kenya. Lastly, drawing from this chapter on human capital externalities and other chapters of this work[99] on private returns to human capital, as a first approximation of social returns, I examine the social returns of human capital in Kenya. No known studies in recent times (in the past twenty years) have provided estimates of social returns for Kenya.

In the remainder of this chapter, I present a brief Data subsection in subsection 4.2, I then turn to discuss results in subsection 4.3. I present a summary and concluding remarks in subsections 4.4.

[99] In the last two chapters (see Oledibe (2023a;2023b)), at the individual level, I examine the effects of education on skill; and the wage effects of education and skill in urban Kenya.

4.2 Data and Methods

4.2.1 Data and Summary Statistics

This study uses the World Bank's STEP dataset for Kenya (Wave 2) fielded between the 1st of August to the 30th of November 2013. This household survey is part of the STEP Measurement program, an initiative to obtain internationally comparable datasets on broad measures of human capital (inclusive of skills and schooling) in developing contexts. Besides, the household surveys, the programme includes employers' surveys, hence, obtaining data on both the demand and supply of skills in labour markets of developing contexts. Implementing the standardised STEP Household Surveys aid in eliciting household and individual level data across the districts of eight provinces in urban Kenya, collecting data on earnings, and levels of schooling, both in actual years of schooling (qualifications) and employment of adults aged,15 – 64, inclusive. The survey instruments include a Direct Reading Literacy Assessment of cognitive skills (in plausible values), designed by the Educational Testing Services (ETS) that also generated the Plausible Values for the extended assessments. Within the Background Questionnaire (the second instrument, besides the reading literacy assessment) is the indirect (self-response) measures of cognitive skills such as reading, writing and numeracy. The background questionnaire developed by the World Bank includes all other modules, such as self-responses on household demographics, dwelling characteristics, health, employment, training, education, personality, behaviour, preferences, language, and family background. As a unit (or observation), a respondent is randomly selected among eligible household members and the (2-2.5) hour survey is administered by the Etude Economique Conseil Inc (EEC Canada).

The selection process is designed and carefully monitored (for compliance) during fieldwork, by the STEP team. The sampling is a 3-stage design stratified by four geographies based on the capital cities and the number of households across cities (see the Stratum Variable in the Data Section). With a response rate of 91.8%, the sample size for urban Kenya is 3894 households and 2355 respondents are in paid employment at the time of the survey. The survey excludes itinerants (see classifications in the 2009 Population Census in Kenya), the unstable, and the war-marred regions of Kenya. The Survey Methodologist oversaw the weighting of the survey sample. The design and implementation of the survey show that it only captures data from cities and urban Kenya, hence, conclusions from the analysis suggest the data set is not deemed representative of Kenya with substantial rural settlements, but rather, conclusions relate to urban Kenya. The dataset was produced (made available) on the 14th of March 2016, by the Development Economics Data Group affiliated to the World Bank. It is freely available in the World Bank Microdata Library.

The STEP survey is one of the very few (if any other) available datasets fielded in non-OECDs having its lineage to the PIAAC for the OECDs, however, it is cross-sectional. Although the non-OECDs still lack useful longitudinal datasets, the provision of the STEP data means similar research carried out in OECDs can now be done for the non-OECDs and results compared across the technology-rich and High-Income OECDs or developed contexts; and the low- and mid-income (or developing) contexts. The STEP data has been used for return estimates; education/skills mismatches; and social inequality and mobility. In this study, I have not restricted the age categories. I present results for the full analytical sample – the employed – but have estimates of employment categories. This subsampling supports a close examine the effects of education and skills as measures of human capital. The focus is on urban Kenya makes it possible to carry out more structured analyses that exploit the specifics of the educational systems (and structural changes within). Table 4.1 provides detailed summary statistics of average educational attainment (years of schooling), average skill (Plausible Values of Reading Assessments) and average hourly earnings in U.S. Dollars, across all districts of the eight provinces of urban Kenya.

4.2.1.1 Specifying the District-Level Variables

Table 4-1 Summary statistics, the STEP Sample and Subsample (the Paid)

Districts	Code	**Summary Statistics for Urban Kenya**					**Summary Statistics for the Paid in Urban Kenya**					
		Freq	**%**	Cum Freq	Average Years of Schooling	Average Skill (Reading Proficiency) _PV	Freq	**%**	Cum Freq	Average Years of Schooling	Average Skill (Reading Proficiency) _PV	Average Hourly Earnings (USD)
BONDO	1	15	0.39	0.39	9***	179.1***	10	0.42	0.42	9.900***	204.8***	2.936***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
BUNGOMA E.	2	15	0.39	0.77	10.40***	195.6***	8	0.34	0.76	10.63***	210.6***	2.048**
					(0.000)	(0.000)				(0.000)	(0.000)	(0.001)
BUNGOMA N.	3	15	0.39	1.16	8.733***	183.3***	9	0.38	1.15	8.778***	182.7***	2.599***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
BUNGOMA S.	4	30	0.77	1.93	8.767***	181.4***	20	0.85	2	9.842***	193.5***	3.330***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
BUSIA	5	15	0.39	2.31	7.400***	175.7***	9	0.38	2.38	8.125***	191.5***	2.677
					(0.000)	(0.000)				(0.000)	(0.000)	(0.093)
ELDORET E.	6	44	1.13	3.44	11.53***	190.8***	28	1.19	3.57	12.48***	198.0***	17.68
					(0.000)	(0.000)				(0.000)	(0.000)	(0.086)
ELDORET W.	7	105	2.7	6.14	10.95***	167.1***	66	2.8	6.37	11.17***	163.2***	3.003***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
EMBU	8	15	0.39	6.52	10.27***	185.2***	11	0.47	6.84	10***	188.0***	2.604**
					(0.000)	(0.000)				(0.000)	(0.000)	(0.004)
GITHUNGURI	10	15	0.39	6.91	6.933***	100.8***	8	0.34	7.18	7.500***	109.1**	0.993***
					(0.000)	(0.000)				(0.000)	(0.003)	(0.000)
GUCHA S.	11	15	0.39	7.29	8.400***	161.9***	9	0.38	7.56	8.111***	164.9***	2.185***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
HOMABAY	12	15	0.39	7.68	9.533***	158.0***	10	0.42	7.98	10***	167.1***	3.936***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.001)
IGEMBE	13	15	0.39	8.06	8.267***	169.2***	11	0.47	8.45	9.600***	194.5***	1.942**
					(0.000)	(0.000)				(0.000)	(0.000)	(0.008)

IMENTI N.	15	30	0.77	8.83	7.667***	167.1***	24	1.02	9.47	7.762***	164.7***	2.785***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
ISIOLO	16	30	0.77	9.6	7.867***	185.6***	18	0.76	10.23	8.467***	204.5***	2.552***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
KAJIADO C.	17	30	0.77	10.37	10.97***	154.7***	19	0.81	11.04	11.44***	152.3***	1.875***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
KAJIADO N.	18	195	5.01	15.38	11.22***	188.5***	117	4.97	16.01	11.28***	182.8***	4.384***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
KAKAMEGA C.	19	29	0.74	16.13	7.517***	166.3***	17	0.72	16.73	9.250***	182.6***	2.806***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
KAKAMEGA E.	20	15	0.39	16.51	8.733***	182.9***	14	0.59	17.32	9.583***	191.8***	3.596
					(0.000)	(0.000)				(0.000)	(0.000)	(0.056)
KALOLENI	21	15	0.39	16.9	9.067***	189.5***	8	0.34	17.66	8.714***	189.6***	8.055
					(0.000)	(0.000)				(0.000)	(0.000)	(0.264)
KANGUNDO	22	15	0.39	17.28	9***	148.8***	7	0.3	17.96	9.200***	142.6**	2.967
					(0.000)	(0.000)				(0.000)	(0.005)	(0.068)
KEIYO	23	15	0.39	17.67	5.800***	125.7***	10	0.42	18.39	6.100***	125.0***	1.519***
					(0.000)	(0.000)				(0.001)	(0.000)	(0.000)
KERICHO	24	15	0.39	18.05	7.267***	179.4***	10	0.42	18.81	8.444***	183.5***	2.176*
					(0.000)	(0.000)				(0.000)	(0.000)	(0.012)
KIAMBU	25	89	2.29	20.34	12***	182.3***	44	1.87	20.68	11.84***	167.1***	3.397***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
KIBWEZI	26	14	0.36	20.7	7.929***	159.7***	10	0.42	21.1	6.333**	132.8***	1.736***
					(0.000)	(0.000)				(0.004)	(0.000)	(0.000)
KIKUYU	27	75	1.93	22.62	12.21***	196.0***	50	2.12	23.23	12.91***	201.2***	4.740***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
KILIFI	28	73	1.87	24.5	11.93***	204.3***	40	1.7	24.93	12.85***	199.3***	6.171***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
KILINDINI	29	180	4.62	29.12	11.03***	186.4***	115	4.88	29.81	11.20***	189.2***	5.878***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.001)
KIPKELION	30	15	0.39	29.51	8.467***	184.7***	11	0.47	30.28	8***	175.4***	1.645***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
KIRINYAGA	31	30	0.77	30.28	12.17***	232.3***	18	0.76	31.04	12.83***	220.2***	6.849***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
KISII C.	32	30	0.77	31.05	9.833***	164.8***	18	0.76	31.8	10.17***	174.9***	2.974***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
KISUMU E.	33	179	4.6	35.64	11.56***	189.3***	102	4.33	36.14	11.98***	187.4***	3.665***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
KITUI	34	15	0.39	36.03	7.267***	147.3***	9	0.38	36.52	6.556***	131.5***	2.645***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
KOIBATEK	35	15	0.39	36.41	8.333***	182.7***	12	0.51	37.03	8.700***	195.5***	1.768***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.001)
KWALE	36	15	0.39	36.8	6.333***	159.4***	11	0.47	37.49	6***	168.5***	1.577***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
LAIKIPIA E.	37	15	0.39	37.19	7.333***	176.2***	9	0.38	37.88	8.333***	190.7***	2.114***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
LAIKIPIA W.	38	14	0.36	37.54	8***	189.0***	11	0.47	38.34	8.818***	203.0***	2.454***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
LIMURU	39	30	0.77	38.32	10.77***	147.1***	21	0.89	39.24	11.05***	145.2***	3.785*
					(0.000)	(0.000)				(0.000)	(0.000)	(0.014)
LOITOKITOK	40	15	0.39	38.7	11.40***	206.6***	6	0.25	39.49	12.17***	200.2***	1.182***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
MAARA	41	15	0.39	39.09	7.600***	165.8***	9	0.38	39.87	8.250***	161.8***	2.854***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.001)
MACHAKOS	42	134	3.44	42.53	9.716***	151.2***	84	3.57	43.44	9.630***	147.4***	1.769***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
MALINDI	43	90	2.31	44.84	10.82***	167.1***	51	2.17	45.61	11.10***	164.9***	3.627***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.001)
MARSABIT	46	15	0.39	45.22	8.800***	177.8***	9	0.38	45.99	8.667***	190.8***	3.298***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
MIGORI	47	30	0.77	45.99	8.400***	158.8***	17	0.72	46.71	8.313***	192.9***	2.116***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
MOLO	48	29	0.74	46.74	9.931***	177.5***	23	0.98	47.69	9.957***	176.5***	1.808***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
MOMBASA	49	195	5.01	51.75	10.40***	156.0***	107	4.54	52.23	10.54***	155.4***	2.233***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
MSAMBWENI	50	30	0.77	52.52	8.767***	194.1***	21	0.89	53.12	8.474***	185.6***	3.153***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
MUMIAS	51	15	0.39	52.9	9***	190.4***	8	0.34	53.46	9.125***	189.4***	1.673***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
MURANGA N.	52	30	0.77	53.67	6.833***	149.2***	24	1.02	54.48	6.583***	145.5***	2.789***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
MWINGI	53	15	0.39	54.06	9.133***	180.6***	6	0.25	54.73	8.667***	188.1***	1.857*
					(0.000)	(0.000)				(0.000)	(0.000)	(0.014)
NAIROBI E.	54	373	9.58	63.64	10.73***	169.4***	229	9.72	64.46	11.26***	168.6***	3.828***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)

NAIROBI N.	55	329	8.45	72.09	11.15***	177.2***	201	8.54	72.99	11.48***	181.8***	4.945***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
NAIROBI W.	56	194	4.98	77.07	9.932***	154.0***	111	4.71	77.71	10.42***	153.1***	2.503***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
NAIVASHA	57	60	1.54	78.61	10.72***	189.8***	36	1.53	79.24	11.29***	203.7***	2.938***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
NAKURU	58	119	3.06	81.66	11.06***	206.7***	76	3.23	82.46	11.51***	211.1***	3.077***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
NAKURU N.	59	29	0.74	82.41	8.448***	117.9***	16	0.68	83.14	7.938***	88.29***	2.115*
					(0.000)	(0.000)				(0.000)	(0.000)	(0.044)
NANDI E.	60	15	0.39	82.79	8.533***	186.9***	12	0.51	83.65	8.909***	204.1***	1.630***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
NAROK N.	61	14	0.36	83.15	8.429***	177.4***	10	0.42	84.08	9.333***	189.6***	2.050***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
NYANDARUA N.	62	15	0.39	83.54	13.20***	236.6***	13	0.55	84.63	13***	225.6***	3.424***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
NYANDO	63	30	0.77	84.31	11.80***	167.9***	13	0.55	85.18	12.33***	157.4***	2.148***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.001)
NYERI N.	64	30	0.77	85.08	15.17***	250.8***	20	0.85	86.03	15.33***	251.7***	11.39*
					(0.000)	(0.000)				(0.000)	(0.000)	(0.031)
NYERI S.	65	30	0.77	85.85	15***	251.8***	21	0.89	86.92	15.30***	246.3***	5.934***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
RONGO	66	15	0.39	86.24	9.733***	197.9***	10	0.42	87.35	11.10***	210.3***	5.514***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
RUIRU	67	105	2.7	88.93	10.15***	170.0***	49	2.08	89.43	10.38***	170.1***	2.058***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
SIAYA	68	15	0.39	89.32	6.467***	134.8***	7	0.3	89.72	7.143***	169.5***	4.088**
					(0.000)	(0.000)				(0.000)	(0.000)	(0.003)
SUBA	69	15	0.39	89.7	8.933***	168.4***	9	0.38	90.11	8***	168.5***	1.949***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
TAITA	70	15	0.39	90.09	9.467***	193.7***	8	0.34	90.45	8.429***	181.5***	2.678*
					(0.000)	(0.000)				(0.000)	(0.000)	(0.030)
TANA RIVER	71	15	0.39	90.47	6.933***	153.5***	11	0.47	90.91	7.545***	163.3***	1.203***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
TESO S.	72	15	0.39	90.86	7.533***	179.2***	12	0.51	91.42	7.900***	189.9***	3.896**
					(0.000)	(0.000)				(0.000)	(0.000)	(0.006)
THIKA W.	73	75	1.93	92.78	10.15***	158.9***	44	1.87	93.29	10.95***	163.7***	3.242***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
TRANS MARA	74	14	0.36	93.14	4.786***	119.8***	11	0.47	93.76	5.444**	132.9***	2.378**
					(0.000)	(0.000)				(0.002)	(0.000)	(0.002)
TRANS NZOIA W	75	59	1.52	94.66	8.525***	131.8***	28	1.19	94.95	8.741***	116.4***	2.927***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
TURKANA C.	76	15	0.39	95.04	6***	102.5***	6	0.25	95.2	4.333	92.64*	1.133
					(0.000)	(0.000)				(0.059)	(0.020)	(0.108)
TURKANA N.	77	14	0.36	95.40	6.714***	122.4***	1	0.04	95.24	13	197.0346	0.000
					(0.000)	(0.000)						
VIHIGA	78	15	0.39	95.79	8.533***	202.3***	9	0.38	95.63	8.778***	204.0***	1.210***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
WARENG	79	59	1.52	97.3	10.47***	179.0***	32	1.36	96.99	9.821***	165.6***	9.676
					(0.000)	(0.000)				(0.000)	(0.000)	(0.069)
WEST POKOT	80	15	0.39	97.69	7.200***	109.1***	12	0.51	97.49	8.167***	129.0***	2.222**
					(0.000)	(0.000)				(0.000)	(0.000)	(0.001)
WESTLANDS	81	75	1.93	99.61	10.45***	173.8***	49	2.08	99.58	11.24***	174.7***	3.024***
					(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
YATTA	82	15	0.39	100	9.933***	134.5***	10	0.42	100	10.10***	154.4***	1.958**
					(0.000)	(0.000)				(0.000)	(0.000)	(0.003)
		3894	**100**				**2355**	**100**				

Notes. The seventy-eight (78) districts are across the eight provinces of Kenya. The sample is deemed representative of urban Kenya. Columns show frequencies, percentages, mean years of schooling (actual), and mean of cognitive skill as reading proficiency (not standardised) for the entire sample and the paid, with mean of hourly earnings (in USD). Besides these averages specific to this study, all other variables used are as defined in previous chapters (see Oledibe, 2023a;2023b). Also, please see the Data Appendix for a full description of all variables used in this study.

The Stratum Variable

The districts in Table 4.1 fall into four (4) strata based on the sample design. The analytical sample for Kenya is stratified by the following geographical areas and based on the number of households (HH) as thus:

Table 4-2 The Stratum

				Mean of the Sample				Mean for the Paid		
Stratum	*City Size/Characteristics*	Freq	%	Avg YoS	Avg Skill	Freq	%	Avg YoS	Avg Skill	Avg H _Earn
1	***Nairobi***	971	24.94	10.69***	169.3***	590	25.05	11.18***	170.8***	3.908***
				(0.000)	(0.000)			(0.000)	(0.000)	(0.000)
2	***Large Cities – Other Large Cities with over 100,000HH***	1001	25.71	10.73***	175.0***	589	25.01	11.03***	176.4***	3.459***
				(0.000)	(0.000)			(0.000)	(0.000)	(0.000)
3	***Medium Cities – 60,000 to 100,000HH***	999	25.65	10.96***	177.4***	583	24.76	11.17***	173.1***	4.455***
				(0.000)	(0.000)			(0.000)	(0.000)	(0.000)
4	***Other Urban Areas***	923	23.7	8.793***	174.8***	593	25.18	9.134***	181.4***	3.131***
				(0.000)	(0.000)			(0.000)	(0.000)	(0.000)
		3894	100	3875		2,355	100	2225		

Notes: Columns show frequencies, percentages, average skill (reading proficiency), average schooling (number of years of schooling) and average earnings (hourly earnings in USD), across strata in Kenya. Source: Author's elaboration of the STEP Household Survey for Kenya.

Firstly, I create dummy variables to capture the effects of the characteristics of each of the Strata, using (4: Other Urban Areas) as a reference category. It is important to note that each of these variables captures the effects of the broad characteristics of the respective strata—including the size of districts, earnings and aggregate human capital.

Stratum_N is a dummy variable indicating a district is in Nairobi Strata

Stratum_L is a dummy variable indicating a district is in Other Large Strata.

Stratum_M is a dummy variable indicating a district is in a Medium Strata.

Stratum_S is a dummy variable indicating a district is in the Small Strata (this makes the reference category).

In this analysis, inspired by the studies of Moretti (2006); Liu (2007) and Acemoglu and Angrist (2000), I create variables used to capture the (wage) effects of aggregate human capital as aggregate schooling (number of years of schooling); and aggregate skills (reading proficiency) across districts. Araki (2020) used a variant of these, to capture the effects of rising educational attainment and skills diffusion (skill proliferation) in OECD countries. Moretti (2006) used the percentage of college-educated (instead) in the US as a measure that gives the externality of schooling, across districts. Liu (2007) used both average schooling and the percentage of college-educated in China to capture the effects of externalities of schooling across cities of China following the approach of Acemoglu and Angrist (2000) (known to use average schooling across states of the United States of America) and the approach of Acemoglu (2006). For Kenya, I use the average number of years of schooling, *Avg_Yos*; and average skills, *Avg_Skill* as average reading proficiency across districts. These make continuous variables that capture externalities of human capital across all districts.

Table 4-3: Summary Statistics, the District-Level Variables

Variable	Brief Description	Observations	Mean	Std Dev	Min	Max
Avg_Yos	***Average Number of Years of Schooling in District***	3,301	10.366	1.498	4.786	15.167
Avg_Skill	***Average Reading Proficiency in District***	3,301	174.590	22.363	100.821	251.772
Stratum_N	***Nairobi Strata***	3,301	0.250	0.433	0	1
Stratum_L	***Other Large Strata***	3,301	0.260	0.439	0	1
Stratum_M	***Medium Strata***	3,301	3,301	0.261	0.439	0
Stratum_S	***Small Strata***	3,301	0.230	0.421	0	1

Note: Source—Author's elaboration of World Bank's STEP data for Urban Kenya

4.2.2 Methods

This chapter summarises the findings from Chapters 2 and 3. Please, refer to the Methods subsections of Chapters 2 and 3. However, in lieu of the individual-level schooling and skills in Chapters 2 and 3, the focus is on the effects of average schooling (Avg_Yos) and average skill (Avg_Skill) across the districts of Kenya.

For Pecuniary externalities of human capital, please, refer to the baseline and 2SLS-IV models specified in Chapter 3 (specifically, refer to Equations 3.2 - 3.4). For non-pecuniary returns to schooling, please, refer to Equation 2.6 specified in Chapter 2.

4.3 Results and Discussions

This (Results and Discussions) subsection will proceed, thus: I start with pecuniary externality, discussing the pecuniary externality of schooling, then pecuniary externality of skill as human capital externalities. I then discuss the non-pecuniary externality of schooling.

4.3.1 Estimates (OLS) of External Returns to Human Capital

Pecuniary Externality of Schooling

Table 3.8 (see Table in Chapter 3) presents OLS (Baseline) estimates of the pecuniary external returns to schooling.

The baseline evidence suggests being in Other Large districts (with over 100 000HHs) besides Nairobi Districts most adversely impacts the hourly wage of individuals. This explains a loss of up to 38.6% of hourly earnings in Other Large districts; a loss of up to 26.9% in hourly wage in the medium districts; and up to 16.1% loss in districts in the Nairobi stratum. Suggesting being in the Small Districts (reference category) has a positive effect on the hourly wage of the individual, this may be a case of excessive supply relative to the demand for labour in larger districts relative to smaller districts. However, compared to Other Large districts, evidence suggests (relatively) minimal adverse wage effect of the substantially larger districts of Nairobi. This indicates the peculiarity of Nairobi, in that, it is the capital of Kenya, where several other factors may be drivers of wages. Hence, the wage effect of the Nairobi stratum may not be in concordance with the other strata.

I now turn to examine external returns. External return to schooling enters as a quadratic term. The effect is negative but increasing (or becoming less negative) with an increase in average schooling across districts (U-shaped). Hence, districts with substantially high average schooling may have positive or substantially less negative external returns to schooling compared to districts with low average schooling. Specifically, the evidence suggests, private return to schooling is positive ranging from 10.2% for an additional year of schooling statistically significant at the 0.1% level. However, with current average schooling across districts for the employed of 10.3 years (Table 3.6), the evidence shows that an additional year of average schooling explains an external return of 0.1106% rise in hourly wage — (calculation: –0.403+2x0.0201x10.3=0.01106)—this entails taking the first order partial derivative, with respect to average schooling. The coefficients of external returns to schooling are statistically significant at the 1% level (at least) and suggest Kenya can further invest in schooling as an approximation of social return of 0.1106%+10.2% = 10.3106%. This finding is inconsistent with the wage effects of district sizes (stratum) as it suggests higher supply (increased schooling, with current demand) favourably impacts wages. The findings from this study are consistent with the study of Kimenyi et al., (2006) that used the 1994 Welfare Monitoring Surveys. However, Kimenyi et al., (2006) found linear and positive OLS estimates of the externality of schooling in Kenya. The findings in this study suggest average schooling in a district must exceed a certain threshold, here (0.403/2x0.0201 = 10.02 years), to have a positive wage effect, otherwise, it will result in a negative externality. This finding is inconsistent with the second proposition of Pritchett (2001) which suggests that rising human capital adversely impacts returns.

Pecuniary Externality of Skill

Still, on pecuniary externality, I now turn to examine skill as an externality of human capital both as a basis to test the robustness of the OLS estimates of the externality of schooling which also presents a basis to improve understanding of the effects of externalities of skill on growth amidst related argument of the lack of skills for growth in developing contexts. Table 3.17 presents similar specifications for average skill as estimated for average schooling in Table 3.8 and the following points are worth noting. The findings suggest that, relative to small districts, being in Other Large districts, adversely impacts hourly wage and this is, at least, a 15.6% loss in hourly wage. Interestingly, no other strata (district sizes) accounted for, have statistically significant wage effects. Suggesting that, being in the Other Large districts has substantial adverse wage effects, relative to small districts, overall, this is consistent when schooling (in lieu of skills) is accounted for as a measure of human capital.

I now turn to the externality of skills. Average skill enters linearly, and effects are positive and statistically significant at 0.1% with a unit rise (in Plausible Value (PV)) of average skill explaining a rise in the individual hourly wage in the range of 0.46-0.55%. Taking the choice specification (without controlling for non-cognitive skills due to multicollinearity), see column (9) of Table 3.17. The externality of skill (0.533%) is at least, greater than the private return to skill which explains a 0.364% rise in the individual hourly wage. I now summarise and draw inferences from these findings for externalities of schooling and skills. Tables 3.8 and 3.17 for externalities of schooling and skill suggest useful consistency in outputs regardless of measures (average schooling or skill) accounted for. In addition to this, the stratum-specific (size) effects show useful consistency, and the private pecuniary returns to education and skill are positive and precisely estimated. However, the pecuniary human capital externalities using average schooling and skill differ substantially suggesting distinctive effects/implications of schooling and skill for growth. Specifically, the external return to schooling has a quadratic (u-shaped) wage effect where average schooling must reach a certain threshold to stay positive, whereas the external return to skill is linear and positive, hence, it has a favourable wage effect regardless of the average skill. Secondly, the externality of skill is substantially greater than the private returns to skill. However, the reverse is the case for the effects of average and individual schooling. Overall, the externality of skill and schooling suggests that the current demand for schooling (and skills) outstrips the supply of schooling (and skill). Furthermore, this is indicative of a useful social return to human capital, hence, the government of Kenya may not relent in its efforts to continue to raise schooling and skill in urban Kenya. Whilst the effects of schooling and skill are (overall) consistent, this may suggest schooling can be taken as skilling. However, understanding the linear and favourable externality of skill is greater than the private return to skill; and the quadratic and favourable externality of schooling is less than the private returns to schooling, this suggests the far-reaching effects of skill should better inform investment in skilling through schooling that require a certain threshold in average schooling to yield positive wage effects. The distinctions in the effects of schooling and skills indicate that, in lieu of a single measure, both measures of schooling and skills should be considered in examining human capital stock.

Schooling does not necessarily mean skilling, especially in developing contexts where many studies argue this (see Pritchett 2001; Hanushek et al., 2015). On one hand, a mechanism that may (at least, in part) explain this distinct wage effect of average schooling and average skill is the complementarity across categories of the educated; and the complementarity across categories of the skilled. For the externality of schooling, the evidence of the favourable wage effects is conditional on a certain level of average schooling, across the district. Hence, this suggests that rising average education (above the 10.02 years threshold) and hence the rising supply of the highly educated, would raise wages or marginal product of the educated in districts with at least 10.02 years of average schooling. This suggests that the highly educated and low educated are not perfect substitutes, particularly in districts where average schooling is at least 10.02 years. Oledibe (2023a, forthcoming) finds that the respondents with post-primary schooling (with an average of 10.2 years of schooling) have higher reading proficiency than those with other higher categories of schooling, in urban Kenya. This suggests that those with high credential categories may not be equipped (or have the required skillset) to perform functions that those with low credentials would normally perform, which is quite unusual. However, for the externality of skills, the complementarity between the low and highly skilled suggests both categories are not perfect substitutes, hence, a rise in average skill or increase in the supply of the skilled raises the wages or the marginal product of the skilled. On the other hand, the consistency in the wage effects of schooling (with over 10.02 years of average schooling across districts) and skill is attributable to a case where a rise in average schooling does result in a rise in average skills as evidence suggests schooling, hence average schooling is not excessive in supply to the point where they result in adverse wage effects. Hence, the demand for schooling and skills raises the marginal product of respondents in urban Kenya. This is inconsistent with the arguments of Pritchett (2001).

In summary, this suggests that access to more schooling (above 10.02 years of average schooling) and skills not only raise wages or marginal product. Those with relatively low education/skill may be more favourably impacted by rising average schooling and skill through their exposure to the additional spillover effects that accrue from being more positioned to receive higher increases in their schooling and skill, over time. This makes the argument for a focus on the rise in average schooling and skill paramount for economic growth in developing contexts, particularly in urban Kenya. I now examine the non-pecuniary externality of schooling, in this study. Hence, the effects of average schooling on individual skills.

Non-Pecuniary Externality of Schooling

Table 2.13 presents OLS (baseline) estimates of the effects of average schooling on individual skills. The evidence suggests that district size adversely impacts adult skills, in urban Kenya. Effects are precisely estimated and statistically significant at the 0.1% level and mean coefficients are greatest in Nairobi districts, the Medium districts before Other Large districts. With a fall in standard skills, in the range of 22-31%, across districts. This suggests that the stratum of small districts explains a rise in standard skills. Overall, findings are consistent with pecuniary returns to district size, where evidence also suggests that the wage effect of the districts with the least number of households, best explains the rise in wages for the employed. I now turn to average schooling. Consistent with the wage effect of average schooling, evidence suggests that average schooling also enters as quadratic terms, negative but increasing (or becoming less negative) as average schooling across districts rises (U-shaped). Particularly, with average schooling of 10.6 years of schooling across districts (regardless of labour market status, see Table 2.1), the evidence suggests an additional year of average schooling across districts explains a standard loss of 2.704% (-

0.256+2x0.0108(10.6)) in cognitive skill. However, an average schooling of ≈12 years of schooling across districts is required to give a non-negative effect of average schooling on skill in urban Kenya. Hence, this requires an additional ≈1.5 years of schooling across districts, to have a positive externality of schooling on skill. To further examine the findings on the OLS estimates of pecuniary human capital externalities (wage effects of average schooling and average skill) and the non-pecuniary human capital externalities (the effects of average schooling on adult reading proficiency), I now turn to the Two-Stage Least Square (2SLS-IV)[100] estimates.

4.3.2 Estimates (2SLS-IV) of External Returns to Human Capital

Pecuniary

Table 3.14 presents outputs of the 2SLS-IV pecuniary estimates of external returns to schooling and skills, whereas the non-pecuniary externality of schooling is found in Table 2.20. Estimates of external returns are obtained using the interaction terms of the policy dummy and the quarter-of-birth in a 2SLS-IV approach. This study acknowledges the limitations of the 2SLS-IV approach (see Appendix A3 for the test of instruments) as taking average schooling and skill as endogenous did not work well. This is not to say the instruments are weak, as they have met some conditions of good instruments for schooling and skill, particularly, the exclusion restriction where the instrument is shown not to influence earnings, except through the endogenous variables; and the instrument results in exogenous variations in schooling and skill, although this is not in all cases. Particularly, the impact of the instrument on skill is taken to be through the effect of the instrument on schooling, hence, the reform sufficiently impacts schooling and schooling in turn, impacts skill. However, the findings on the 2SLS-IV provide a useful basis to examine the OLS estimates for external returns. The latter shows useful robustness, particularly with its quadratic outcome for external return to schooling, and the linear outcome for external return to skill. These baseline (OLS) outcomes for external returns to schooling and skill have strong conceptual (theoretical) underpinnings to the motivation of this study, and I am convinced it is ‘reasonably’ free from bias. The 2SLS-IV instruments are further investigated

[100] Just for emphasis, the preference specification of this study is the OLS specification, the 2SLS-IV specification is deemed to be less robust due to issues with the validity of instruments (please see Appendix A3).

(in future studies) to obtain more consistent instruments that will have positive test (of weak instrument) outcomes.

In Table 3.14, whilst no statistically significant return exists for the pool or main analytical sample, the (substantial) variability in the wage effects of average schooling (external returns) exists across subsamples of the main analytical sample. Particularly, for the males and the private wage employed where external returns are linear, negative, and statistically significant at the 5% level. Although it is a usual practice to exclude individuals of a certain age, such as those thought not to have completed their formal education. However, with the emphasis on economic growth in this study, hence the consideration for net effects of externalities that come with the inclusion of all (or a representative of all) aged 15-64 in the labour force. The samples/subsamples of interest are not limited in this study. Hence, I include all in the labour force, with the age range: 15-64. Furthermore, Table 3.15 and 3.16 show outputs of the first-stage equation. Whilst evidence from the first-stage equation suggests, the reform only impacts (positively) the average schooling of females and quarter of birth substantially impact the average schooling of males, the main analytical sample and the informal. However, the interaction of the reform and first and second quarter of birth only positively impact the average schooling of males. As earlier highlighted, whilst the external return to schooling is statistically insignificant (hence, not different from nil) for the main analytical sample. Evidence suggests, the external returns of -25.3% and -23.2% in hourly earnings of the male gender and those in private-sector wage employment respectively. Hence, a one-year rise in average schooling is explains a loss of 25.3% and 23.2% in hourly earnings for the males and the private-sector wage employed. Table 3.14 also shows that the male gender and private wage employed have positive and statistically significant private returns to their schooling. However, for average and individual skills, evidence suggest external returns and private returns to skill are not different from nil across the main analytical sample and all subsamples of interests. However, in a slightly different 2SLS-IV specification that accounts for tenure which also enters as a quadratic term (see panels C and D of Table 3.14), the evidence suggests the private return to skill and schooling for the male are positive (0.767%) and (25.6%) respectively and statistically significant at the 5% level. In addition to this, in this slightly different specification of the 2SLS-IV, the private returns to schooling for the private wage employed is positive (25.1%) and statistically significant at the 1% level and the external return to schooling is a loss of 22.6% of hourly earnings statistically significant at the 5% level. However, controlling for tenure in the 2SLS-IV specification (that instrumented both individual and average schooling and skill, with the time-varying reform indicator) may suggest that the 2SLS-IV specification is spurious. Hence, I consider the 2SLS-IV specification without the tenure variable (see panels A and B) to be relatively robust.

Comparing the OLS and the 2SLS-IV estimates of pecuniary external return to schooling and skill, the 2SLS-IV evidence suggests whilst the external return to schooling is negative, particularly, for the male and private-sector wage employed, the 2SLS-IV estimates for externalities of skill are generally inconsistent. Hence, interestingly, the OLS estimates, with the current levels of average schooling of the employed at 10.3 are positive, small, and more precisely estimated. However, the OLS estimates for externalities of schooling enter as quadratic terms (which is negative and becoming less negative) with rising average schooling across districts. Specifically, the effects of average schooling of 10.3 years (see Table 3.6) for the employed is not the case of the OLS return for external skill that is positive, linear, and strongly statistically significant. In considering the external returns to schooling for the private-sector wage employed for the baseline (OLS) and 2SLS-IV. The OLS estimates show substantial responsiveness to average schooling across districts. This reflects the substance or true effects of average schooling in line with the predictions (see Introduction—Chapter One) on the schooling-earnings-growth link in urban Kenya. Similarly, comparing OLS and 2SLS-IV estimates of pecuniary external returns of skill gives inconsistent results. Specifically, using similar specifications for schooling and skill, findings suggest across the main analytical sample and other subsamples of interest, whilst the OLS estimates of pecuniary external returns are mainly positive and statistically significant, entering linearly. However, the 2SLS-IV pecuniary external returns are statistically insignificant across the main analytical sample and the subsamples of interest. The bias of the 2SLS-IV appears to be from the defects in the instrument and/or the 2SLS-IV specification that did not account for non-linear terms. The outputs of the first stage equation suggest this, as the instruments have no statistically significant effect on the average schooling and skill across districts; and the individual schooling for these subgroups. Essentially, this means the instrument instrumenting average schooling and skill, particularly for the subgroup/s in question does not result in useful exogenous variation of average schooling and skill which means the current outcomes for the 2SLS-IV estimates are spurious. Ultimately, the baseline (OLS) estimates are sufficiently (and relatively) robust compared to the 2SLS-IV estimates. In addition to this, as earlier argued, the OLS estimates of external returns to schooling and skill are sufficiently robust, accounting for individual schooling, health, and personality traits, among other covariates. Hence, minimising possible multicollinearity or omitted variable biases (see Tables 3.17 and 3.23). Ultimately, the OLS estimates may give useful insights into the Average Treatment Effects (ATE)[101] relative to the 2SLS-IV estimates of average schooling and skill deemed spurious (see Appendix A3) like the 2SLS-IV estimates of the externality of schooling. Hence the baseline (OLS) estimates of the pecuniary externalities of schooling and skill are deemed consistent[102].

Non-pecuniary

I now discuss the non-pecuniary externality of schooling as a possible basis for the outcome of pecuniary external returns to schooling, specifically comparing the OLS to the 2SLS-IV estimates. Table 2.20 presents the 2SLS-IV estimates; Table 2.13 presents the OLS estimates of the effects of aggregate schooling (on individual skill) as a non-pecuniary externality of schooling. The initial evidence suggests a negative non-pecuniary external return to schooling for the—2SLS-IV and OLS specifications. Specifically, a unit increase in average schooling across districts explains a standard loss of 24.5% in individual reading proficiency, statistically significant at the 5% level. However, further analysis suggests this 2SLS-IV estimate of external return to schooling is inconsistent with the OLS estimates. However, like the OLS estimates of (the pecuniary) external returns to schooling, the non-pecuniary external return enters as quadratic having a negative effect (that becomes less negative with rising average schooling across districts) on individual skill. Specifically, the OLS evidence suggests a rise of about 1.5 years in average schooling is required to make the non-pecuniary externality of schooling favourable (see subsection 4.3.1). Interestingly, relative to the 2SLS-IV estimates, evidence suggests the OLS estimates of non-pecuniary external returns are also more consistent and precisely estimated (lower standard errors) compared with the outputs of the 2SLS-IV estimations. This, among others as previously highlighted (please, see outcomes of the tests of instruments, in Appendix A3) affirms the OLS pecuniary and non-pecuniary external returns to schooling are sufficiently robust and give useful approximation to the Average Treatment Effects.

Schooling or Skills as Measures of Human Capital

In the previous subsection (4.3.1) on the analysis of estimates based on OLS, it was concluded that in developing contexts, a consideration for both measures of human capital

---

[101] As known the OLS estimator gives the Best Linear Unbiased Estimator with the seven OLS conditions in place.

[102] However, it is good to note that the 2SLS-IV estimates of private returns to schooling and skills are reasonably consistent showing positive wage effects of schooling and skills. The 2SLS-IV estimates for individual schooling and skill are the choice output in this study.

externalities—average schooling and average skill—is crucial for a useful understanding of human capital and their externalities. The examination of the 2SLS-IV relative to the OLS estimates did not suggest otherwise.

The distinct (OLS) estimates of external returns to schooling and skills (with the former being negative and becoming less negative with the rise in schooling (quadratic); and the latter being linear and positive), suggest that both measures of average schooling and skills are distinct, having distinct effects on individual earnings. Particularly, the positive effect of the externality of skills suggests it is an appropriate measure of human capital stock. The negative and quadratic external returns to schooling suggest reasonable care should be in place in interpreting this externality of schooling as a measure of the human capital stock linked to productivity. This is the effect of the externality of schooling which can become positive with substantially increased average schooling. Overall, relative to the externality of schooling, the minimal positive and strongly statistically significant effect of the externality of skill seems to be better linked to the minimal productivity of the human capital of the country or region. Fundamentally, this small mean effect but strongly statistically significant and positive externality of skill may also suggest the effects of relatively weak (or lack of) skills from formal schooling in a developing context (like urban Kenya) as Pritchett (2001) suggests (this is further discussed in the Conclusion to thesis in Chapter 6, please, see interconnected themes). It was noted that, primarily, skill provides a useful basis for capturing human capital. However, this evidence shows that at a certain level of aggregate schooling, effects on skills and earnings change drastically (from adverse to favourable). This compels the need to estimate human capital (and its externalities) with individual and aggregate schooling and skill. The consideration or use of both measures—individual and aggregate schooling and skill—to capture human capital and human capital externalities is of crucial policy relevance, especially in a developing context.

4.4 Summary and Concluding Remarks

Using the World Bank STEP data, this study examines estimates of pecuniary and non-pecuniary external returns to measures of human capital stock, in districts across urban Kenya.

Effects of Stratum (District Size) on Individual Schooling and Skill

The descriptive evidence suggests substantial variation in average schooling and skills across districts (see Table 4.1). The baseline (OLS) estimates suggest external returns to schooling and skill are substantially subsumed by the effects of the size[103] or the number of households in districts (strata). See columns (6) (7) (8) (14) (15) and (16) of Table 2.1 where accounting for average schooling and/or skill in a regression containing the strata variables gives spurious effects. The effects of the district size on individual wages and cognitive skills, although, this varies based on the strata of the district. However, effects are consistent, in that, the effects of the district size (strata) on wages are almost the same (in nature not necessarily in magnitude) as the effect of the district size (strata) on skill—see Tables 2.1, 3.8 and 3.17. This provides useful evidence suggesting skill is a mechanism through which human capital impacts earnings. Specifically, with the Small Stratum (districts with under 60 000HHs) as a reference category, evidence from the baseline (OLS) estimates suggests, that being in districts within the Other Large Stratum (over 100 000HHs) explains an average of 38.6% loss in hourly earnings. Similarly, being in the Other Large Stratum also explains an average of 25.9% standard loss in cognitive skills. Furthermore, being in the Nairobi Stratum (over 100 000HHs) explains an average of 16.1% loss in hourly earnings, further evidence suggests a 30% standard loss in cognitive skills is attributable to residing in the Nairobi Stratum. Being in districts of the Medium Stratum (with 60 000<no of HHs<100 000) explains an average of 26.9% loss in hourly earnings and an average standard loss of 28.7% in cognitive skills. Overall, the evidence suggests that being in districts of the Small Stratum with under 60 000HHs has more favourable (or less adverse) effects on individual skills and hourly earnings. I now summarise findings on external returns that give insights into social returns. For external returns, the preference specification is on the baseline (OLS) estimates. However, the 2SLS-IV estimates support an assessment of the robustness of OLS estimates.

[103] The size of districts in urban Kenya is based on the stratum of the district, following the three-stage stratified sampling procedure of the STEP data. Please, see the data section for details of the strata of the districts.

Pecuniary External Returns to Schooling and Skill

The 2SLS-IV approach shows no statistically significant external returns to schooling in urban Kenya, using a representative sample of the employed, in urban Kenya. However, the evidence indicates that a one-year rise in average schooling results in a loss of 25.3% and 23.2% in the hourly earnings of the males and the private-sector wage-employed, respectively. Comparing this to quadratic outcomes for the OLS estimates with average schooling of 10.3 years for the employed, 10.8 for the males and 10.8 years for the private sector wage employed[104]. I now turn to calculate the corresponding OLS estimates for external returns to schooling for the employed, the male and private-sector wage employed respectively, as thus[105]:

- 41% + 2(2.04) % x 10.3 =

+1.024%

- 35.6% + 2(1.81) % x 10.8 =

+3.496%

- 50.6% + 2(2.44) % x 10.8 =

+ 2.104%

Doing these indicates the OLS specification shows a positive pecuniary externality of schooling with the current levels of average schooling across the categories considered. Comparing this to the 2SLS-IV estimates that are negative, it is indicative that the estimates from the former suggest a more robust outcome amidst the limitations (of the latter), cited earlier. As earlier highlighted (see Introduction), a positive estimate of social return is argued to be a fair basis for involvement and investment in human capital by the government of a

[104] Please see Table 3.6 for the descriptive evidence—for average schooling across categories of genders and the employed. Table 3.9 is for external and private returns for the respective categories of gender and the employed.

[105] This is the first-order partial derivative of average schooling outcome in the regression output (quadratic) in terms of the respective average schooling across the district.

country. The sum of external and private returns is deemed a first approximation of (the pecuniary) social returns. Hence, using the OLS estimates, corresponding estimates of the pecuniary social returns for the employed (regardless of the category of the employed), the male and the private-sector wage employed is[106]:

10.7% +1.024%=

11.724%

9.86% +3.496%=

13.356%

10.7% + 2.104%=

12.804%

This favourable externality and private returns to schooling, result in favourable (pecuniary) social returns for the male and private-sector wage employed. These suggest evidence of the government's under-investment in schooling, in urban Kenya. I now examine the pecuniary externality of schooling, for the rest of the gender and employment categories, to understand categories of gender or employment with the most/least externality that will best aid understanding of how government under- or over-investment in schooling impacts all employment or gender categories. Using the simple calculations of the pecuniary externalities and social returns as earlier[107]:

Female: - 45.5% + 2(2.26) % x 9.9 = **- 0.752%** + 10.6% = 9.9%

Informal: - 36.3% + 2(1.88) % x 9.6 = **- 0.204% +** 6.9% = 6.7%

Formal: - 50.2% + 2(2.11) % x 13.4 = **6.34% +** 15.9% = 22.24%

Starting with the female gender and the informally employed, the evidence suggests, a negative pecuniary externality of schooling of **-0.752%, and -0.204%** respectively. Whilst

---

[106] This simply entails adding the corresponding pecuniary private returns to the external returns.

[107] Please, see Table 3.6 for average schooling; and Table 3.9 for private and external returns.

this ultimately results in a positive pecuniary social return as they have positive pecuniary private returns, raising the average schooling of the female gender (currently at 9.9 years of schooling) and the informally employed (currently at 9.6 years of schooling) to the threshold of 10.02 years of schooling as earlier discussed[108] will result in a non-negative outcome of pecuniary externality of schooling. However, the formally employed have positive pecuniary externality of schooling and positive social returns suggesting the government is not yet overinvesting in schooling across all categories of gender and employment discussed so far.

Interestingly, turning to the public service wage employed and those in entrepreneurship, the findings suggest no statistically significant pecuniary externalities of schooling exist for these (even with average schooling well over the 10.02 years threshold). However, these categories of the employed have positive and statistically significant pecuniary private returns. To further examine the pecuniary externalities of schooling for the public service wage employed and those in entrepreneurship, I turn to the pecuniary returns to skill. The evidence (in Table 3.18) suggests that the pecuniary private and external returns to skill are statistically insignificant for the public sector wage employed and those in entrepreneurship. Hence, not only is the nil pecuniary externality of schooling consistent with the effect of the pecuniary externality of skill, but the pecuniary private returns to schooling and skill are inconsistent for the public service wage employed and those in entrepreneurship. This suggests the pecuniary social returns to the skills of the public-service wage employed and those in entrepreneurship are nil. This strongly suggests evidence of over-investment in skill in these categories of the employed. These employment categories are the most important for driving economic growth, particularly, determining growth of other employment categories. This presents a fundamental issue with the use (or application) of skills, as opposed to a lack of skills, in socially productive activities. This is very much in line with the argument of Prichett (2001) that suggests schooling yields little or no skill; schooling is privately remunerative but socially unproductive. This is particularly upheld as schooling has substantial pecuniary private returns but no pecuniary external return. The overall picture is made clear with nil pecuniary social returns to skills that stem nil pecuniary private and external returns to skill. This is a case of misplacement of skill—where cognitive skills are not applied economically in private and socially productive activities—these have substantial implications for growth. This problem requires urgent attention and one way to handle this is by useful employment or work policy that considers (and emphasises) skill

[108] The turning point is where the quadratic outcome of externality of schooling has a nil log-hourly-earnings outcome.

beyond mere schooling, with useful Entrepreneurship Education and Training programmes. The outcome of the use of skill and schooling shows support for arguments that suggest both measures are useful in assessing human capital, in developing contexts. This is made clear in the case of the public service wage employed and those in entrepreneurship. Beyond these categories, the pecuniary private and externality of schooling and skill are consistent across other categories of the employed. Suggesting the pecuniary private, external, and social returns make it clear the government of Kenya can further invest in schooling and skill.

Non-Pecuniary External Returns to Schooling

I now turn to summarise and give concluding remarks on the findings of non-pecuniary external and private returns[109]. Irrespective of genders and employment categories (the pool, not the employed), the evidence from the 2SLS-IV estimates (see Table 2.20) of the non-pecuniary external returns suggests, a one-year rise in average schooling results in a standard loss of 24.5% points in cognitive skills and a one-year rise in individual schooling results in a standard rise of 31.8% points in cognitive skills of the individual. Suggesting favourable private return of individual schooling on individual skill; and adverse externality – effect of aggregate schooling on individual skill. This ultimately indicates a positive non-pecuniary social return to schooling (netting external and private returns) indicating support for more investment in schooling. Hence, a case of underinvestment in schooling. However, heterogeneity in the non-pecuniary external and private returns suggests the disadvantaged, particularly, those that have fathers and/or mothers without post-secondary education have negative social returns to their schooling—a case where the adverse externality of schooling is weightier than the favourable private returns to their schooling—at first, this may suggest overinvestment in their schooling or a case where additional funding will not be supported. But turning to OLS estimates of non-pecuniary private and external returns to schooling, the evidence suggests such negative externality that led to negative social return is attributable to the substantially low schooling of the disadvantaged. I now turn to the OLS estimates—see Tables 2.13 and 2.1. The evidence suggests from the pool (column 6 of Table 2.13; Table 2.1) with average schooling of 10.6 an additional year of schooling for an individual explains an average of 13.6% standard rise in cognitive skill; however, for an additional year of average schooling to result in a non-negative non-pecuniary externality, the respondent must have attained about 11.85 (-0.256/(2x0.0108)) years of schooling (≈12). This means a 1.5-

[109] non-pecuniary external and private returns which entail the extent to which average schooling, and individual schooling impact individual skill.

year increase in average schooling is required to have a non-negative external return. The subsample of the 'disadvantaged' may have lower educational attainment, compared to the current average of 10.6%, suggesting a higher rise in schooling may be required to attain a non-negative external return. The outcome of the 2SLS-IV estimate appears quite consistent resulting in a higher positive private return and a lower negative external return. This is true for the pool of both the OLS and the 2SLS-IV specifications. The higher negative outcome for external returns for the disadvantaged in the 2SLS-IV estimates shows useful consistency with the more robust OLS estimates where the effects of the non-pecuniary externality of schooling are negative and expected to be weightier for the disadvantaged that may require a higher rise in average schooling for a non-negative external return. Although the OLS specification is deemed robust, the 2SLS-IV estimate is a useful comparator in this study. Particularly, for the OLS specification, the external return for the pool (see column (6) of Table 2.13) is a loss of 2.704% (-0.256+2x0.0108(10.6)) of standardised reading proficiency for an additional year average schooling above 10.6 years. From the previous subsection, an increase of ≈1.5 years of average schooling, should yield a non-negative skill outcome. This outcome is regardless of the status of the respondent. It is understood that the disadvantaged would require more school relative to the pool or the advantaged as indicated by the higher negative externality for the disadvantaged (relative to the pool) in the 2SLS-IV specification. To demonstrate this. In Table 2.14 the OLS specification for the pool requires 32.9/ (2x1.47) = 11.2 years of average schooling for a non-negative externality, however, column (5) suggests the disadvantaged (having a father without post-secondary schooling) suggest to have a non-negative externality of schooling an average of 28.7/(2x1.26) = 11.4 years of schooling. However, whilst the 2SLS-IV specification is deemed robust, the choice of the OLS is for consistency and simplicity, in that it captures or communicates more insights with the quadratic outcome for externality of schooling. Secondly, the weakness of the instruments (see Appendix A3) in the 2SLS-IV approach is another reason for the choice of the OLS specification. However, just to be clear that both specifications are not expected to give the same mean effects as the OLS is at best an indication of the Average Treatment Effect and the 2SLS-IV estimate gives the Local Average Treatment Effect, that is specific to those impacted by the reform. Overall, the findings indicate a negative non-pecuniary externality and a positive social return, requiring more investment in schooling, particularly in raising access for higher average schooling across districts in urban Kenya. This is certainly not a case of over- but under-investment in schooling for skill in urban Kenya.

Put together, as possible basis of the poor growth of developing countries Pritchett (2001) third proposition suggests little or no skill from schooling; and the first proposition suggests cognitive skills are privately remunerative but socially unproductive. This is not exactly the case in the context of urban Kenya. The consistency in the pecuniary and non-pecuniary external and private returns suggests substantial evidence of skills from schooling (particularly for the disadvantaged), which further explains earnings. The fact that negative externality becomes less negative (or effectively, becomes positive) with rising average schooling, across districts accentuates not only individual schooling but aggregate schooling, substantially explains skill and this, in turn, impacts earnings. However, this comes with some exceptions—the public service wage employed and those in entrepreneurship—where schooling and skill yield little or no externalities. The evidence suggests that schooling is indeed privately remunerative and socially unproductive as private returns to schooling are substantially high, but externalities are not different from nil. This is regardless of the threshold or level of schooling attained. Ultimately, efforts to raise schooling raise skills and earnings in urban Kenya. This study adds to the literature by providing useful estimates of the private, social, and external returns. In addition to these, this study also improves understanding of Pritchett's (2001) propositions, providing evidence from urban Kenya. An important policy insight (offered by this study) is the requirement to further invest in average schooling for skill or useful productivity in Kenya. Further policy contributions drawn relate to the emphasis on skill in public service wage employed and those in entrepreneurship. Finally, a need for Entrepreneurship Education and Training is emphasised. These insights are crucial for policymaking on growth and development in urban Kenya. This study strongly emphasises the need to use schooling and skills as measures of human capital, particularly, in developing contexts.

5 Returns to Education and Skills in a Dynamic Framework

5.1 Introduction

I now introduce a dynamic approach to obtain useful estimates of returns to education and skills in Kenya.

5.1.1 Why Technology of Skills Formation?

It is well documented that educational attainment, skills and labour market outcomes exhibit useful relationships. However, the effects of these relationships vary across regions of the world. In the literature, robust empirical evidence (studies) deploys longitudinal datasets, ideally with some panel features. It is rare to find such datasets in developing (relative to developed) contexts. Hence, using (the available) single-period cross-section of data in the developing contexts can inhibit the robustness (particularly, at the extensive margins) of such research evidence. More specifically, using a single cross-section of data as in Chapters 2, 3 and 4 of this study presents an additional econometric requirement to test the robustness of the evidence reported. Hence, this compels deploying a dynamic framework that can account for this data limitation to synthesise robust estimates of the effects of schooling on skill, and the effect of skill on earnings.

Besides issues of data limitation, the central place of skills as an outcome of schooling and predictor of economic productivity including earnings (and other labour market outcomes) is crucial for policymaking. Of late, Kenya, a developing country in sub-Sahara Africa has continued its campaign for competency-based schooling stemming from its labour market needs (see Introduction in Chapter 2). Following the human capital framework (Becker 1964; Mincer, 1974) estimates of the effects of schooling on skills (Chapter 2); and estimates of the wage effects of skills (Chapter 3) have been examined in reduced forms. Whilst it is argued that accounting for possible biases in each of these individual estimations (as in Chapters 2 and 3) gives reasonably robust estimates, particularly at the intensive margins, a suitable dynamic framework of structural relations where each of the individual relations (chapter 2—effects of schooling on skill; and chapter 3—effects of skills on earnings) are simultaneously estimated as a system[110], may further test the robustness of the existing findings (see chapters 2 and 3).

Specifically, findings of chapters 2 and 3 that suggest re-estimations in a dynamic framework to improve estimates in the intensive margins include the following:

The returns to a year of schooling for those with the ISCED2 credential category remain greater than those with the ISCED34A credential category—the ISCED34A is a level higher than the ISCED2 credential category. Interestingly, the evidence from Chapter 2 where skill from schooling is examined suggests that schooling substantially impacts cognitive skills. However, inefficiently. The inefficiency in schooling here means that, relative to higher credential categories, schooling most impacts skills for those with lower secondary education (ISCED2), which suggests that, relative to other credential categories, education is most productive in the ISCED2 category of schooling. Whilst this is consistent with the return estimates of ISCED2 greater than ISCED34A, the understanding of human capital theory (where productivity in human capital comes through skill) means one would expect this category (with ISCED2 credential category) of the employed to be most productive or have highest returns since it yields more skill relative to other credential categories. Hence, arguably, the effects of signalling for higher credential categories (credentialism) in the labour market may mean that the return to schooling for the employed with lower secondary credential category (ISCED2) appears inconsistent with their skill level (or productivity) in the light of the human capital theory that 'rightly' suggests schooling impacts earnings through productivity or skill. However, this evidence of the relatively low wage effects of a schooling category that explains the highest skill in Kenya is inconsistent with the understanding of human capital theory that suggests the effect of schooling (or human capital) on earnings is through skill in the form of increased productivity. Such inconsistencies in estimates are not new. A plethora of studies argue that schooling gives little or no skill in developing countries, particularly, in sub-Saharan Africa (see Pritchett 2001; Hanushek and Woessmann, 2008). These studies attribute this phenomenon to the poor quality of schooling in the region, with this, one would also expect that not only is the relationship between schooling and skill impacted by the quality of schooling, but by extension, the relationship between skill and earnings may differ substantially as many factors including labour market characteristics are involved. As a response to these quality issues, a strand of the literature has focussed on estimating 'quality-adjusted' or 'quality-

[110] A system of structural and measurement relations, accounting for issues of endogeneity of schooling and data limitation by treating skills and labour market outcomes (such as earnings) as latent variables.

consistent' return estimates (see Hanushek and Zhang, 2006). Hence, regardless of the region of the world, whilst the link across earnings, schooling and skills remains complex, this is at least in part due to the nature of skill formation—it is tough to measure skills possessed. Hence, examining and establishing the extent to which schooling yields skill, and in turn, the extent to which skills explain labour market outcomes simultaneously, accounting for the latency of skill and other outcomes of interest is crucial for useful insights that overcome the shortcomings of the previous analyses, in line with the human capital framework, in reduced form. Hence, effectively this entails a further examination of the relationship across earnings, schooling, and skill for the employed in urban Kenya in structural relations, accounting for the latency of skills and other outcomes of interest.

A dynamic framework can be far-reaching in this study, particularly, further testing the testable predictions of Chapter 2—the effects of background characteristics. In Chapter 2, a quasi-experiment is carried out to examine the effects of background characteristics (parental education and low socioeconomic status) on the schooling and skill of the ward. The evidence from the analysis suggests that background characteristics impact schooling but not the skills of the ward. I argue that the outcome of the quasi-experimental (Difference-in-Differences) analysis is partly due to a limitation in the measure of skill, besides the effects of the ATET estimator. The use of a cross-section of data and the not easily observable nature of skill is a major limitation of the outcome of the analysis. Unlike the measures of schooling, skill is deemed latent by its nature. Hence, an attempt to fully observe skill with a mere cross-section of data may bias estimates involving such relations. Hence, the 'unobservable' nature of skill, and the substantial data limitation should strongly justify treating skill as a latent variable. Whilst the framework used in the previous chapters does not treat skills as latent variables, a more dynamic framework that allows such and addresses this limitation may unravel new insights and further test the robustness of the evidence (or estimates) from previous analyses.

Hence, to overcome data limitations; and other possible (inherent) biases of reduced-form models that relate to the (aforementioned) relationships, inspired by the work of Krishnakumar and Nogales (2020), amidst the limitations[111] of this dynamic framework, I

[111] To do this, I turn to a framework that (comprehensively) accounts for skill formation. Hence, I deploy structural equations of the Technology of Skills Formation—was explored by Cunha and Heckman (2007). This aids an examination of skill from schooling; and the wage effects of skill simultaneously. A growing number of studies have explored the Technology of Skills formation, particularly in developed contexts (for a few studies please, see those of Cunha, Heckman, and Schennach, 2010; and Heckman, Humphries, and Veramendi, 2018). The framework demands substantial data, covering a wide range of information that

deploy an adapted version of the Technology of Skill Formation framework. Using a similar dataset (a single period cross-section) —with a few variables that capture past information—in Kenya, as the study of Krishnakumar and Nogales (2020) in Bolivia. This dynamic framework—the adapted Technology of Skills Formation Framework—attempts to resolve the data and econometric challenges discussed so far, as it not only allows treating skills and the related labour market outcomes (inclusive of earnings) as latent variables but further accounts for data limitations by using some variables with past information of respondents. In addition to this, it addresses the endogeneity of schooling. The modelling specification makes it possible to estimate simultaneously, the effects of background characteristics on schooling and skills (Chapter 2); the effect of schooling on skills (Chapter 2); and the effects of skills on earnings (Chapter 3). In summary, the utility of this approach is that it tests the robustness of estimates at the intensive and extensive margins which provide a basis to (further) test the testable predictions (of Chapters 2 and 3) in this study.

5.1.2 The Objective: Research Question

The objective of this study is to deploy a dynamic framework—the Technology of Skills Formation—which accounts for data limitations and the inherent biases of reduced-form models. Hence, the objective is to obtain a more consistent estimate of returns to schooling, skills, and background characteristics in urban Kenya.

5.1.3 Reviews of the Related Literature and Contributions

The review of the literature is around the use of structural modelling to provide consistent estimates of the Earnings from Skills; Skills from Schooling; and Schooling and Skills attributable to Background Characteristics—parental education and socioeconomic status.

As earlier highlighted, the study of Cunha et al., (2010) in the OECDs sets the scene for this study, the Technology of Skill Formation framework (see Cunha and Heckman, 2007). Cunha et al., (2010) present, an examination of the effect of schooling on labour market outcomes, accounting for the dynamic nature of skill formation. The framework has been

---

spans a significant lifetime of the respondents, from childhood to adulthood, with information/data requirements covering (but not limited to) parental education and inputs on their ward's schooling. This sort of dataset is hard to come by in non-OECDs. Hence, the suitability of this approach for studies in regions with data limitations presents an additional challenge. It is argued that the data requirement involved in operationalising such models is a major inhibitor of the use of such models in research in the non-OECDs (see Laajaj and Macours (2017)). However, the study of Krishnakumar and Nogales (2020) overcame this data limitation using useful variables and a suitably adapted version of the technology skill formation framework.

applied extensively[112] in OECD contexts. However, there is understandably limited evidence of the application of the Technology of skills formation in non-OECD contexts. Whilst a few studies like those of Villa (2017) and Sanchez (2017), in the non-OECDs, have used longitudinal datasets, and exploited the Technology of Skill Formation approach, they have considered health variables in their studies. However, most empirical works in the non-OECDs only use (the available) cross-sectional datasets. However, the recent study of Krishnakumar and Nogales (2020) contributed substantially to this literature, overcoming data constraints, by using the World Bank STEP Household Survey, a mere cross-section of data, that includes extensive skills measures. At the time of their (Krishnakumar and Nogales, 2020) research, the STEP dataset was the most useful available dataset for such study in Bolivia. In so many aspects, this study draws from (and relates to) the study of Krishnakumar and Nogales (2020). However, I explore the case of Kenya, advancing the frontiers of knowledge by exploring a similar dataset—the World Bank's STEP Household Survey for Kenya—and deploying the same version of the Technology of Skills Formation framework. Hence, doing these makes it possible to present comparable estimates of the returns to background characteristics, schooling and skill. The study of Krishnakumar and Nogales (2020) gave useful consideration[113] to the latency of skills and a measure referred to as work-related well-being. However, beyond dimensions of work-related well-being as measures of 'good job' in their study, in this study, with an interest to further examine 'earnings' in employment; and 'formal/informality' of employment, I define three dimensions of a labour market outcome to include, 'earnings', 'formal employment opportunity' in addition to 'safe work environment'. Hence, in this study, I consider the latency of skills and the market environment, in developing contexts. Hence, in addition to closely observing 'earnings' as a separate dimension in this analysis, the use of firm size, affiliations with social security in employment and signing of contract aid useful examination of the effects of 'formal employment opportunity' as a dimension, as opposed to mere 'employment opportunity' using firm size and contribution to social security, as used in the study Krishnakumar and Nogales (2020).

This study presents the first evidence of the use of the Technology of Skills formation to examine the links: between schooling and skill; skill and labour market outcomes; and

[112] For a few studies, see Heckman et al. (2011; 2018); Brunello and Schlotter (2011).

[113] With the measure of skills having two key dimensions for cognitive and non-cognitive skills, and an additional latent outcome measure referred to as *work-related wellbeing*—deemed to capture several aspects of a good job with three key dimensions for—employment opportunities and earnings; overtime work; and safe work environment.

background characteristics and schooling/skill in sub-Sahara Africa. The study of Heckman et al. (2018) is one of the few related studies that use the same approach, comparing findings to outcomes of the reduced form approach, however, in the OECDs context, specifically, in the United States of America. Heckman et al. (2018) affirm substantial wage effects of education and find that, relative to low-ability individuals, 'high-ability' individuals have substantial pecuniary returns to their schooling above high school, suggesting university or tertiary education may not benefit all depending on abilities/skills/traits. Furthermore, the study of Heckman et al., (2018) finds that the pecuniary returns differ across categories of schooling depending on observed and unobserved characteristics including abilities. These findings relate to the previous analyses in this study where returns differ substantially across credential and skill categories. However, the approach adopted in this study which entails examining simultaneously, the effects of schooling on skill and the effects of skills (cognitive and non-cognitive) on earnings can lead us close to making similar conclusions (as in Heckman et al. (2018) in the United States of America) for Kenya.

The use of the Technology of Skill formation approach in the specified relations of this study also relates and supports further examinations of the claims of other known empirical research like those of Pritchett (2001); and Hanushek et al. (2008) that suggest schooling yields little or no skills in the non-OECDs. In addition to this claim, Pritchett (2001) suggests that the lack of (and the wrong use of) skills from schooling contributes to poor economic growth and development in the region. Examining the effects of schooling on skill, and/or the effects of skills from schooling (as in this study) gives some insights into the arguments of Pritchett (2001) and those of Hanushek et al. (2008), in developing contexts.

Furthermore, the flexibility of the structural equation modelling in this framework makes it possible to examine the effects of background characteristics — such as parental education and socioeconomic status — on the schooling and skill of individuals. This may provide useful insights and effectively support the findings of Chapter 2, hence, effectively adding to the literature on intergenerational transmission of abilities and educational mobility (see Heckman et al. (2006); and Black et al. (2005)).

Finally, understanding the extent to which investment in schooling impacts the formation of skills (cognitive, and non-cognitive) has gained attention in the OECDs (see Carlsson et al., (2015); Carneiro et al., (2007); and Carneiro and Heckman (2003)). Whilst it is understood that non-cognitive skills are more malleable, particularly among school children in OECD contexts (for evidence from the US, see Carneiro and Heckman (2003)), Carlsson et al.,

(2015) show that some aspects of cognitive skills for adults can be raised by short schooling term in the Swedish context. However, this understanding of how schooling impacts cognitive and non-cognitive skills has gained minimal attention in non-OECDs, as little or no evidence exists in sub-Saharan Africa. This is partly attributable to data constraints as earlier mentioned. However, amidst this data constraint, useful innovation in econometrics now makes it possible to overcome some of the challenges presented by the data limitations in the non-OECDs (see Todd and Wolpin, 2003; 2007) and unravel useful insights on the effects of schooling on both cognitive and non-cognitive skills. This makes it possible to provide similar evidence on the malleability of skills in developing contexts with a lack of useful longitudinal data. In the non-OECDs, the study of Krishnakumar and Nogales (2020) in the context of Bolivia sets the scene for this study for Kenya, with useful adaptation of the Technology of Skills formation framework (see Cunha and Heckman 2007) to provide some similar evidence for those that exist in the OECD context (see Carlsson et al., (2015) for Sweden; Carneiro and Heckman (2003) for the US). Somewhat consistent with the OECDs, Krishnakumar and Nogales (2020) find that, whilst above average non-cognitive skill is attainable with primary schooling, such level (above average) of cognitive skill is only attainable with tertiary schooling, in the Bolivian context. Inspired by the approach of Krishnakumar and Nogales (2020), this study examines the effects of schooling on both cognitive and non-cognitive skills in urban Kenya, amidst the data limitation. Whilst it will be useful to contribute more by providing further analysis or examination of the malleability of cognitive and non-cognitive, the scope of this study—aimed at testing the robustness of previous analysis (Chapters 2 and 3) prevents this. However, this is left out and will be a major objective for future studies in urban Kenya.

The rest of this chapter is as follows, in subsection 5.2, I discuss the Model and Variable Specifications. I then present and discuss the results in subsection 5.3 with the Summary and Concluding Remarks in subsection 5.4.

5.2 Model and Variable Specification

5.2.1 Model Specification

5.2.1.1 Introduction: The Analytical Framework

The Technology of Skill Formation framework provides a useful basis for re-testing the testable predictions of this study. This framework provides a system of structural and measurement relations that link and re-estimate each of (the previous relations) (see Chapters 2 and 3) estimated in reduced forms. The first—chapter 2—examines the relationship between educational attainment and skills. The second—chapter 3—examines the relationship between skill and earnings (labour market outcomes). The analytical framework links Chapter 2 to Chapter 3. Hence, it provides a two-step process to demystify the relationship between schooling and labour market outcomes, accentuating the central (moderating) place of skill in this relationship between schooling and labour market outcomes. Hence, this suggests a condition where investment in schooling accrues skills that raise productivity and labour market outcomes. As earlier highlighted, in the introduction, the labour market outcome of interest, although inspired by the work of Krishnakumar and Nogales (2020), however, it is different from the measure used by Krishnakumar and Nogales (2020) in several aspects. The labour market outcome of interest is deemed latent (unobservable), however, measured by three key dimensions of the labour market outcomes that include—earnings, formal employment opportunities and a safe work environment (see Data subsection for more).

The informal labour market of Kenya is argued to warrant accounting for these dimensions of labour market outcomes, as understanding how cognitive and non-cognitive skills explain these is of policy relevance. For skills, one important advantage of this framework of structural and latent relation over the reduced-form analysis is that, whilst it was argued that cognitive and non-cognitive skills may not be estimated simultaneously (see mediation analysis in Chapter 3), the conditions (see structural systems and assumptions) in place, makes this possible in this framework. Hence, accounted for are the two key characteristic skills—cognitive and non-cognitive (or personality traits)—influenced by schooling in different ways and known to have different effects on labour market outcomes. Each of these is deemed latent (unobservable), however, they are observed through several measurable values and traits (see Data subsection of chapters 2 and 3 for more) of each characteristic skill.

5.2.1.2 Skills from Investment in Schooling; and Endogeneity in Schooling

In this framework skill is a mediator between schooling and the measure of labour market outcomes. Chapter 2 emphasises the effects of intergenerational family background characteristics on skills through its effects on educational attainment.

Similarly, in the technology of skill formation, the latency of skills over time makes skill acquisition a dynamic process where intergenerational family background characteristics through nature (genes) and nurture (environment) which includes investment in skill formation (schooling) are pivotal to skill formation, particularly, in childhood and adolescence. Specifically, the technology of skills (each of cognitive and non-cognitive skills) formation for a respondent is a function of the following considerations: Parental endowment (e.g., parental education); the vector of current skill stock (even at birth, there is evidence of skill through nature or genetics passed from parents); investment for skill acquisition (educational investment) in the child.

The technology of skill formation is functionally specified as thus:

$$\theta_{t+1} = f_{t+1}(p, \theta_t, I_t, u_{t+1}), \forall t \leq T \ldots\ldots \quad (1)$$

where $t = 1, 2, .., T$ is the timespan of skill acquisition with T developmental stages.

Hence,

θ_{t+1} denotes the technology of skill formation that makes the stock of cognitive (θ_{t+1}^C), and socioemotional or non-cognitive skills (θ_{t+1}^N) at $t + 1$. Hence, $\theta_{t+1} = (\theta_{t+1}^C, \theta_{t+1}^N)$ at $t + 1$. θ_{t+1} is a function of θ_t, which is the stock of the vector of cognitive and non-cognitive skill stocks at t. Hence, the technology of skill formation at the t developmental stage is a period of lead, from the t.

(p), is a measure of parental endowment or characteristics which is time-invariant and can include parental socioeconomic status, skill, and educational attainment.

Finally, u_{t+1}, captures other observables (this can include shocks and work experience) and the unobservables that impact skill formation.

(I_t), is a measure of investment for skill formation, this can be captured by educational attainment which is mainly deemed parental effort but can very much be the respondent's effort as they grow older. In this study, of much consideration is the government effort or interventions in the process of the investment for skill formation, captured explicitly with variables that are exogenous to the system involving interaction terms of the reform dummy (p1985_) and each of the quarters of birth (Q2, Q3 and Q4). Please, see Chapter 2 for a lengthy discussion on government intervention; and Chapter 3 for the use and tests of these instruments. Note that, f_{t+1} is strictly increasing varying in time and concave, with respect to I_t, and it is twice continuously differentiable (see Cunha and Heckman, 2007).

Dynamic complementarity is defined by $\partial^2 f_{t+1}(p, \theta_t, I_t, u_{t+1})/\partial\theta_t \partial I_t' > 0$; Self-productivity is conditioned on $\partial f_{t+1}(p, \theta_t, I_t, u_{t+1})/\partial\theta_t > 0$; Whilst the former (dynamic complementarity) describes a case where the skill stock acquired by t-1, θ_t makes the resultant investment in at t, I_t more productive. This explains why returns to schooling investment are higher at later developmental stages of the lifecycle for children or individuals with higher initial skill stock, θ_t. The latter (self-productivity) describes a case where higher skill formation is attributable to high skill in the preceding period. For the disadvantaged, the combined effects of dynamic complementarity and self-productivity explain the relatively high returns to educational investment for the young relative to the adolescents (Cunha and Heckman, 2007). Generating useful insights on skill formation using the structural equation (1) would ideally require longitudinal data as opposed to a mere cross-section of data. Doing this best unravels the dynamics of skill over its developmental phases in the dynamic framework (of the model) that accounts for this. As mentioned earlier, the research in the developing contexts is inhibited (relatively) by useful longitudinal datasets. However, inspired by the study of Krishnakumar and Nogales (2020), by substitution, I operationalise a recursive resolution of the skill function, substituting successively for skills in past developmental stages in (1), and (2) as thus:

The latency or unobservable nature of skill makes it possible to normalise the skill stock at the initial timespan to 1 whilst maintaining the generalisation.

$$\theta_{t+1} = g_{1,t+1}(p, I_t, I_{t-1}, I_{t-2}, \ldots, I_1, w_{1,t+1}) \ldots\ldots\ldots (2)$$

Here, $g_{1,t+1}$ is as defined; and $w_{1,t+1}$ is the resulting error term which includes all errors, $u_{t+1}, u_t, u_{t-1}, \ldots$

Using (2), neither self-productivity nor dynamic complementarity are defined using this specification as θ_t, the skill in the previous developmental stage is eliminated in (2). However, (2) makes it possible for all other features of the framework (technology of skill formation) to be preserved providing more utility in a developing context that is deficient in longitudinal data. (2) only requires data on previous investments for skill acquisitions (schooling, in this case); and only a period of observed skill, deemed to be the skill with one period of lead, which are easily obtained from cross-sectional data.

For (2) to provide plausible estimates of the effects of the investments in schooling $(I)\ on\ skill\ (\theta)$, the orthogonality condition which suggests omitted or unobserved variables (within the error term) are uncorrelated with I, must be upheld. However, current specification, particularly, assumptions (innovations) in place to arrive at (2) from (1) stem from data limitation which means the error term may contain unobserved variables correlated with I (omitted variable bias), meaning orthogonality conditions will seldom be in place. Other possible issues that may introduce bias in the estimate of I is typical of functions with schooling variables that may have been specified with error (measurement error bias); and another is that I may be explained by (latent) skill (reverse causality bias). The studies of Heckman et al. (2011); and Cunha et.al (2010) on the technology of skill framework made clear biases in the estimate of I_t, attributable to omitted variables and reverse causality as earlier defined. Both biases suggest the endogeneity of I_t, meaning the orthogonality condition where $w_{1,t+1}$ is uncorrelated with I_t is not preserved. Hence, results in the estimate of I_t in (2).

To address the bias in the estimate of $I_{t,}$ attributable to the endogeneity of I_t in (2) Krishnakumar and Nogales (2020) specified an 'investment policy function' inspired by Cunha et al. (2010) that proposed dividing the error term, $w_{1,t+1}$ into two parts, with a part, σ_{t+1} that is uncorrelated with all independent variables of (2) where the orthogonality condition is preserved. The other part, τ_{t+1}, make variables that are unobserved, correlated or partly identical to I_t, however, orthogonal (iid) to other independent variables in (2)

Hence, if $w_{1,t+1} = (\sigma_{t+1}, \tau_{t+1})$, from (2), (2) is rewritten as the following:

$$\theta_{t+1} = g_{1,t+1}(p, I_t, I_{t-1}, I_{t-2}, \dots, I_1, \sigma_{t+1}, \tau_{t+1}), \forall t \leq T \dots\dots\dots\dots (3).$$

Modifying (1), the investment policy function is expressed as thus:

$I_t = m_t(p, \theta_t, r_t, \tau_t), \ \forall t \leq T$………………………………………………………... (4).

Where r_t instruments I_t in (3) and makes variables which may only explain θ_{t+1} through effects on I_t. Inspired by the study of Krishnakumar and Nogales (2020) in obtaining (2) from (1) by substituting successively for past skill (θ_t) mitigating the data limitation, I repeat the same, substituting θ_t in (4) by (3) modifying (4) giving (5) as thus:

$I_t = m_t(p, I_{t-1}, \dots, I_1, r_t, \overline{w}_t)$………………………………………………………… (5)

Here, all variables are as defined with $\overline{w}_t$ as the error term. I now turn to model earnings from skill (please, see analytical framework).

5.2.1.3 Earnings from Skill (as Investment in Schooling); and the Latency of Skill and Earnings

From (1) where the timespan of skill acquisition, $t = 1, 2, \dots, T$ is divided into T development stages. T is deemed when an individual has acquired the skills or stock of human capital (after initial investments in skill), θ_T, required so they can start to function and benefit from the labour market. Hence, at $T + 1$, the stock of human capital or resources available include, θ_{T+1}, which is defined by their stock of cognitive and non-cognitive skills; and other resources and circumstances represented by ($\bar{S}_{T+1}$) that can explain labour market outcomes. Hence, if Y represents the labour market outcomes, Y_{T+1} defines the relationship that explains how human capital relates to labour market outcomes.

$Y_{T+1} = g_2(\theta_{T+1}, \bar{S}_{T+1}, w_{2,T+1})$…………………………………………………….. (6)

All terms in (6) are as earlier defined; and $w_{2,t+1}$ is the error term.

The latency of skills suggests that skills are ever evolving and may at best be measured, assessed, or observed in models with other variables (see Kautz et al., 2014; Heckman et al., 2006). S$o\ far, at\ t + 1, \theta_{t+1}$ has been observed as the stock of skill. However, if at $t + 1$, Z^{θ}_{t+1} denotes the observable stock of skill (observed + unobserved = observable skill). Both skill types (cognitive and non-cognitive) may be presented as thus: $Z^{\theta}_{t+1} \equiv (Z^{C}_{t+1}, Z^{NC}_{t+1})$. The latency of skill entails a relationship $Z^{\theta}_{t+1}, \theta_{t+1}\ and,\ \alpha_{1,t+1}$ where the latter is the unobservable elements that impact the indicators of skills and the process of skill transformation, b_1(.) to the observable indicators.

Hence given as thus:

$$Z_{t+1}^{\theta} = b_1(\theta_{t+1}, \alpha_{1,t+1}) \ldots\ldots (7)$$

The defined measure of labour market outcome is a latent (a theoretical construct), hence, not a directly observed measure[114]. However, measured by earnings and a set of labour market outcomes in all three dimensions of the outcome (earnings, formal employment opportunity and measures of safe work). Hence, with the latency of the labour market outcome, Y_{T+1}, similarly, the measurement equation is given as (8) and $b_2(.)$ is the process of the transformation of the labour market outcome to the observed dimensions (earnings, formal employment opportunity and safe work environment) and $\alpha_{2,t+1}$ make the unobservable factors in the process.

$$Z_{t+1}^{Y} = b_2(Y_{T+1}, \alpha_{2,t+1}) \ldots\ldots (8)$$

5.2.1.4 Estimation and Identification Strategy – The Structural Systems (SEMs)

Having a cross-section of data on background characteristics, investment for skill formation (schooling or educational attainment) in the past (t); to obtain estimates of the effects of investment and background characteristics on earnings and skill in adulthood $(T+1)$. Following the approach of Krishnakumar and Nogales (2020), I re-model each of the structural relations, (2)(6)(5)(7) and (8) at $(T+1)$ as a system of simultaneous equation model (9) as thus: with the first, a structural relation relating skill to educational investment (2); the second, a structural relation relating earnings to skill stock (6); the third, a structural relation relating educational investment to past educational investments, parental skill (education) and socioeconomic status (5); and the fourth, a measurement relation, showing the latency of skill by, relating observable skills to observed skill and an error term (7); and the fifth, a measurement relation, relating observable earnings to observed earnings as thus:

[114] In urban Kenya, about eighty percent (80%) of the workforce are informally employed (based on the STEP sampling), with this and the issue of reticence in reporting earnings for survey data (see Azfar and Murrell, 2009) and the associated complexities (corruption, informality of employment and multiple sources of unaccounted and unreported income from larger black economies) in sub-Sahara Africa, I make a case for treating labour market outcomes as latent. Specifying a measurement relation with dimensions that include earnings, formality of employment and safe work. The latter is inspired by the work of Krishnakumar and Nogales (2020) for Bolivia.

$$\theta_{T+1} = g_{1,T+1}(p, I_T, I_{T-1}, I_{T-2}, \ldots, I_1, w_{1,T+1})$$

$$Y_{T+1} = g_2(\theta_{T+1}, \bar{S}_{T+1}, w_{2,T+1})$$

$$I_t = m_t(p, I_{t-1}, \ldots, I_1, r_t, \overline{w}_t).\ With\ I_0\ fixed; for\ 1 \leq t \leq T$$

$$Z^{\theta}_{T+1} = b_1(\theta_{T+1}, \alpha_{1,T+1})$$

$$Z^{Y}_{T+1} = b_2(Y_{T+1}, \alpha_{2,T+1})$$

(System 9).

Put together, the structural relations make a system of Simultaneous Equation Model (SEM, with latent variables) inspired by Muthen (1983; 1984); and explored by Krishnakumar and Nogales (2020). In the previous analyses (see Oledibe 2023a; 2023b (forthcoming)) each of the first three models ((2), (6), and (5)) are operationalised in reduced forms (without structures or additional assumptions). However, operationalising these simultaneously with the structures in place gains richness of the dynamics of skill formation accounted for (by the framework), in explaining related outcomes (skills and earnings). I now turn to discuss transforming (System 9) to give (System 10), as a structural system of Simultaneous Equation Model (SEM) responsive to the set objectives and acknowledging the specifics (e.g., dataset) or context (Kenya) of this study.

As earlier discussed, with t = 1, 2…, indicating the timespan of skill formation involving T skill development stages. At t = 1, this is the first stage in skill formation that depicts the early childhood or preschool stage. At t = 2, the second stage in skill formation is taken to be the formal schooling stage, hence, primary to tertiary schooling. The third and final stage in skill formation is the period in which the survey was fielded which makes, $T + 1$ as earlier defined. Therefore, so far, with T = 2, $T + 1 = 3$ shows that, $T + 1$ is the third skill formation stage.

Hence, (System 10) is given as thus:

$$\theta_3 = g_1(p, I_1, I_2, w_{1,3})$$

$$Y_3 = g_2(\theta_3, \bar{S}_3, w_{2,3})$$

$$I_2 = m_2(p, I_1, r_2, \overline{w}_2)$$

$$I_1 = m_1(p, r_1, \overline{w}_1)$$

$$Z_3^{\theta} = b_1(\theta_3, \alpha_{1,3})$$

$$Z_3^{Y} = b_2(Y_3, \alpha_{2,3})$$

(System 10).

Here, the measurement and structural relations make a System that accounts for the endogeneity in schooling, in a framework that combines the observed respondent's schooling (I_1, I_2) deemed endogenous, parental characteristics (p) deemed exogenous, exogenous resources that impact earnings $(\bar{S}_3)$, vector of skill (Z_3^{θ}) and labour market outcomes (Z_3^{Y}) are deemed endogenous, as well as the socioeconomic variables that impact schooling (r_1, r_2) are deemed exogenous; and unobserved skill stocks (θ_3) and earnings (Y_3) in adulthood, are deemed endogenous and the error terms in each of the structural and measurement relations $(w_{1,3}, w_{2,3}, \overline{w}_2, \overline{w}_1, \alpha_{1,3}, \alpha_{2,3})$, that make the system. The endogeneity concerns of these variables call for an identification strategy on the entire system of measurement and structural relations. Following the works of Cunha et al., (2010); Cunha and Heckman (2008); Krishnakumar and Nagar (2007); Skrondal and Rabe-Hesketh (2004); and Muthen (1983; and 1984), the study Krishnakumar and Nogales (2020) proposes a strategy that attempts to secure consistent parameter estimates. I follow a similar approach in this study, as thus:

For ease of exposition, starting with the structural relations, where $\mu = (I_1, I_2, \theta_3, Y_3)$, as vectors of the observed and latent structural variables sized (p1 + p2 + 2 + m) × 1. Hence, with the observed investment (I_1, I_2) variables as a p1 – dimensional vector variable of the observed educational investment in the pre-school period, I_1; p2 – dimensional vector variable of the observed educational investment in the formal schooling period, I_2. (θ_3) is a vector of 2-dimensional sub-vector of the latent non-cognitive (θ_3^{NC}) and cognitive (θ_3^{C}) skills; finally, Y_3 is an m-dimensional sub-vector of earnings deemed partially observed. Hence, if p1 + p2 = p, with $X = (p,\ \bar{S}_3, r_1, r_2)'$ as a k×1 vector of exogenous variables (controls) which include, current circumstances that impact earnings $(\bar{S}_3)$, schooling investments (r_1, r_2); and parental characteristics (p).Therefore, if the number of elements in $r_1, r_2, \bar{S}_3, and\ p$ are represented by $\iota_{1,}\iota_{2,}\kappa_{1,}\kappa_{2,}\ respectively, then \quad \iota_{1,} + \iota_{2,} + \kappa_{1,} + \kappa_{2,} =$ k.

As usual, for most empirical applications of the framework, the typical first condition for identification entails a proposition of additive linear forms in the structural relations (as in System 10) that are separable in the error terms as thus: $w_{1,3}, w_{2,3}, \bar{w}_2\ and\ \bar{w}_1$, for the respective functions as thus: $g_1(.), g_2(.), m_1(.)$ and $m_2(.)$, with these assumptions in place and having C, and D as coefficient matrices; and w as a vector of all errors in the structural relations (as in System 10). Then, for the ith observation, the structural relations between each of the individual elements of μ; as well as elements of X, is given as thus:

$$C\mu - DX - w = 0 \quad \text{(11)}$$

Hence, here $E(w) = 0;$ and the full variance-covariance matrix $V(w) = \Sigma$

The exclusion restriction is usually the second identification condition. I now turn to discuss how this is operationalised in the structural relations (of System 10). For the earnings (Y_3) and skills (θ_3) relations, inspired by the approach of Krishnakumar and Nogales (2020) that follow the idea of the technology of skill formation, I allow skills, θ_3 (cognitive and non-cognitive) to directly impact earnings, (Y_3) excluding educational investments (I_1, I_2) from the earnings structural relation. Educational investments (I_1, I_2) are deemed to directly impact skills, from the skill (θ_3) structural relations, hence, educational investments, (I_1, I_2) are deemed to (indirectly) impact earnings through skills (θ_3) in the earnings (Y_3) structural relation. Whilst current circumstances, $\bar{S}_3$ directly impact earnings, (Y_3) but not adult skill formation, (θ_3); parental characteristics, p (including measures of parental skill) are taken to impact adult skills, (θ_3) but not directly in the earnings structural relation (Y_3). The exclusion restrictions for the investment structural relations (I_1, I_2) is reached by instrumenting, I_1 with r_1; and I_2 with r_2. Put together, the implication of these mathematical translation on the theoretical restrictions for $(Y_3, \theta_3, I_1, I_2)$ is seen in the configuration of elements of (11) in (12) with $C_{11}, C_{12}, C_2, C_3, D_1, D_2, D_3\ and\ D_4$ are coefficients of matrices C and D as earlier described and the identity matrix is I.

$$\begin{bmatrix} 1 & 0 & C_{11} & C_{12} \\ C_2 & 1 & 0 & 0 \\ 0 & 0 & 1 & C_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_3 \\ Y_3 \\ I_2 \\ I_1 \end{bmatrix} — \begin{bmatrix} D_1 & 0 & 0 & 0 \\ 0 & D_2 & 0 & 0 \\ 0 & 0 & D_3 & 0 \\ 0 & 0 & 0 & D_4 \end{bmatrix} \begin{bmatrix} p \\ \bar{S}_3 \\ r_2 \\ r_1 \end{bmatrix} — \begin{bmatrix} w_{1,3} \\ w_{2,3} \\ \overline{w}_2 \\ \overline{w}_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

(Subsystem 10.1)

I now turn to the measurement relations of System 10. The identification conditions between the observable skills and earnings (Z_3^{θ}, Z_3^{Y}) and the observed skills and earnings (θ_3, Y_3) draws from similar approaches to structural relations following the studies of Krishnakumar and Nogales (2020). Hence, this entails a proposition of additive linear forms in the measurement relations (as in System 10) that are separable in the error terms as thus: $\alpha_{1,3}\ and\ \alpha_{2,3}$ as error terms for their respective relations as thus: $b_1(.)\ and\ b_2(.)$. From earlier discussions, where $Z_3^{\theta} \equiv (Z_3^{C}, Z_3^{NC})\ is\ the\ observable;\ and\ \theta_3 \equiv (\theta_3^{C},\ \theta_3^{NC})$ is the observed. Each of the cognitive and non-cognitive components makes vectors of the related skills deemed latent and the observed (θ_3) distinguishable from the observable (Z_3^{θ}) makes the basis of the measurement relations in skills. Similarly, as earlier argued, particularly in the context of the non-OECDs, although earnings are largely observed. However, the informality of employment in the black economy where limited tax information makes a measurement relation a useful approach to accounting for the unobserved earnings in the context (see System 10). In addition to informality in employment and the related tax consequences in a black economy, reticence in reporting earnings and related survey instruments is a major source of biases in estimates. Hence, distinguishing the cognitive, non-cognitive and earnings measurement relations to give (Subsystem 13), where ϱ_j, j = $\{C, NC, Y\}$ are deemed factor loading as thus:

$$Z_3^{C} \ = \ \delta_C + \varrho_C \theta_3^{C} + \alpha_{1,3}^{C}$$

$$Z_3^{NC} \ = \ \delta_{NC} + \varrho_{NC} \theta_3^{NC} + \alpha_{1,3}^{NC}$$

$$Z_3^{Y} \ = \ \delta_Y + \varrho_Y Y_3 + \alpha_{1,3}^{Y}$$

(Subsystem 10.2)

Inspired by the study of Krishnakumar and Nogales (2020) that followed the approach of Skrondal and Rabe-Hesketh (2004). If the vectors, $\alpha_{1,3}^{j}$ where j $= \{C, NC, Y\}$ are independent 2x2 with a mean of zero. Measured (observed) indicators in relationship to a common (latent) variable are correlated. With the latent factor as the mere channel for the correlation. Hence, making allowance for the generalised case of the correlation of both observed skills as thus: $\text{cov}(\theta_3^C, \theta_3^{NC}) \neq 0$ to partly account for the complementarity of cognitive and non-cognitive skills given the limitation in data (having just a cross-section of data). Also allowed for is the non-zero correlation between the possible dimensions of observed earnings (Y_3) to obtain the common possible unobserved earnings, this may be through the impacts of unobserved adult circumstances that may also impact the observed earnings (see the structural (Y_3); and measurement relations of Z_3^Y in System 10). Hence, a scaling condition that normalises (to unity) one element for the vectors ϱ_j, j $= \{C, NC, Y\}$ is deemed to apply to Subsystem 13. Based on the studies of Krishnakumar and Ballon (2008); Muthen (1983 and 1984), under these conditions, having at least three indicators for skills and earnings preserves the identification strategy for each of the respective measurement equations (Krishnakumar and Nogales, 2020). From System 10, unlike skills and earnings, there is no measurement relation for investments, this presents some consistency threats where measurement error in the investment variable may be deemed not accounted for (see, Cunha et al. (2010); Krishnakumar and Nogales (2020)). It is important to highlight that this is attributable to data limitations, as there is (only) one observed investment variable across each of the skill development stages considered. However, to mitigate the possible effect of this inconsistency and the (possible) related consequence in the measurement of the investment variable, the investment variable is deemed to be directly observed and endogenous. Hence, inspired by the study of Krishnakumar and Nogales (2020), structural relations modelled, (I_1, I_2) accounts for the endogeneity in the investment variables that are instrumented with measures whose effects on varying outcomes come through the investment variables (see I_1, I_2 in System 10).

After specifying the identification for System 10, Subsystems 12 and 13 with related assumptions/conditions respectively account for the endogeneity in the structural and measurement relations in the System 10. In estimating Subsystems 12 and 13 the maximum likelihood approach by Muthen (1983; 1984) is operationalised (see Krishnakumar and Nogales, 2020). In a vector, for the ith individual, collecting the observed variables (exogenous) as $\boldsymbol{L_i} = (r_1, r_2, p,\ \bar{S}_3)$i with a size of $q_1 x 1$; and the observed variables (endogenous) for the same individual, including the observable indicators of skills and earnings, as $\boldsymbol{P_i} = (Z_3^{\theta}, Z_3^{Y}, I_1, I_2)\boldsymbol{i}$ with a size of $q_2 x 1$ as earlier described. Hence, let the number of observed variables in the system be represented as q = $q_1 + q_2$. For the observed endogenous variable, the conditional moment is given as thus: $E(\boldsymbol{P_i}|\boldsymbol{L_i}) = \lambda$ and $V(\boldsymbol{P_i}|\boldsymbol{L_i}) = \gamma, \forall i$. Where theoretical expressions for λ *and* γ are obtained by substituting the structural relations for the observed skill, θ_3 *and* the earnings, Y_3 into the respective measurement equations. Therefore, λ *and* γ contain all parameters in the system. Hence, if ζ is the column vector of unknown parameters contained in the model inclusive of the variance-covariance elements. Therefore, with an assumption of conditional normality of $\boldsymbol{P_i}$ given $\boldsymbol{L_i}$, $\forall i$. Hence, for an individual i, the log-likelihood is denoted as thus:

$$\text{logHi}\ (\boldsymbol{P_i}|\boldsymbol{L_i}\colon \zeta) = -\frac{1}{2}(\text{q} \times \log(2\pi) + \log(\det(\gamma)) + (\boldsymbol{P_i} - \lambda)'\gamma^{-1}(\boldsymbol{P_i} - \lambda)).$$

Hence,

$$\text{logH}\ (\zeta) = \sum_{i=1}^{N} \text{logHi}\ (\boldsymbol{P_i}|\boldsymbol{L_i}\colon \zeta) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (14)$$

is that (log-likelihood) of the entire system with the omission of the arguments of $\boldsymbol{P_i}$ and $\boldsymbol{L_i}$, $\forall i$. For the sample, where N is the number of individuals. (14) is maximised to give parameter estimates $\hat{\zeta}$. The heteroscedasticity-consistent sandwich formula for a quasi-maximum likelihood estimator. This makes it possible to obtain robust or asymptotic standard errors. By the application of an empirical Bayes application, the use of the parameter estimates makes it possible to draw scores of the skills (θ_3) and earnings Y_3 variables deemed latent (see Krishnakumar and Nogales (2020).

5.2.2 Variable Specification.

5.2.2.1 The STEP Data

This study uses the publicly available STEP household survey of the World Bank for Kenya which is part of the second wave of the survey fielded between August and November 2013. This dataset is representative of the urban Kenya and findings are interpreted as thus. As earlier stated, the World Bank's STEP Household Survey is the first initiative to measure (or elicit) detailed skills of respondents in low- and mid-income countries (also referred to as the non-OECDs). In the non-OECDs, the lack of useful longitudinal observational datasets inhibits useful empirical research as earlier discussed (see Introduction). Although cross-sectional, the wide range of skills in the STEP makes it the only available dataset in Kenya that supports operationalising[115] the explored theoretical framework—technology of skills formation—providing useful empirical analysis underpinned by the theory. Please, see the Appendices and the Data Sections of Chapters 2 and 3 for further details on the STEP data. I now turn to specify each of the variables used in this study.

5.2.2.2 The Variables

Measures of Cognitive and non-Cognitive Skills:

For cognitive skills, rather than using self-reported measures or other measures that may capture some cognitive skills of the respondents, I use a 'more established and comprehensive' indicator of cognitive skills in the literature. I use the ten Plausible Values (PVs) that capture the reading proficiency of respondents, and hence, the (cognitive) skills of the respondents. The test was administered by ETS (Educational Testing Services). Please, see further description and detailed descriptive evidence in Chapter 2 and the Data Section of the Appendix Chapter. For non-cognitive skills, whilst the WB STEP provides several useful measures of non-cognitive skills as averages of batteries of survey instruments that elicit personality or socioemotional traits, to be consistent with the rest of the chapter, I use only the 'established' five personality or socioemotional traits termed the Big 5 taxonomy—this includes, the extraversion average; conscientiousness average; openness average; emotional stability average; and agreeableness average—as measures of non-cognitive skills for consistency with reduced form analyses from previous chapters. However, of interest are the other measures of non-cognitive skills, in the World Bank's STEP dataset for Kenya. Hence, grit average, decision-making average, hostile bias average, risk aversion and time preference are considered in subsequent analysis. Please, see further description and descriptive evidence of the Big 5 in the Data Section of Chapter 3 and the Data Appendix Chapter.

[115] Operationalising the theoretical framework explored comes with some innovations to mitigate the deficiency attributable to the (static) data which presents some defects in maximising the potential of dynamic framework. However, the current set up (see Model Specification) makes sufficient use of the framework (from which useful insights are drawn) amidst the data limitation.

Table 5-1 Summary Statistics, Sample of the Employed

Variable	Brief Description	Obs	Mean	Std. dev.	Min	Max
	Measures of non-cognitive skills (personality traits)					
extraversi~v	Extraversion Average, continuous	2,054	2.8719	0.5808	1	4
conscienti~g	Conscientiousness Average, continuous	2,054	3.2459	0.5083	1.5	4
openness_av	Openness Average, continuous	2,054	3.0081	0.5504	1	4
stability_av	Emotional Stability Average, continuous	2,054	2.7068	0.4984	1	4
agreeablen~v	Agreeableness Average, continuous	2,053	2.8635	0.5626	1	4
	Measures of cognitive skills (plausible values)					
pvlit1		2,063	178.8288	87.6515	0	375.0276
pvlit2		2,063	178.6496	88.9369	0	367.0915
pvlit3		2,063	178.8879	87.231	0	372.5338
pvlit4		2,063	179.3052	87.70861	0	365.2552
pvlit5		2,063	179.5645	86.5329	0	363.2965
pvlit6		2,063	179.6078	86.07631	0	362.5683
pvlit7		2,063	178.24	87.4111	0	361.7662
pvlit8		2,063	178.6546	87.4590	0	387.0413
pvlit9		2,063	179.3061	86.8256	0	358.3947
pvlit10		2,063	180.3605	86.7425	0	349.3105
	Measures of labour market outcomes					
earnings_h~d	Hourly Earnings in USD, continuous	1,902	4.0263	11.0597	0.0310	260.5714
m4c_q20	Firm Size: no of paid workers, categoric	1,351	3.5403	2.0268	1	7
m4c_q21	Job affiliated with social security, categoric	1,358	1.6053	0.4890	1	2
m4c_q18	Have signed a contract, categoric	1,176	2.1701	0.8713	1	3
m5b_q02	Lifted above 50lb at work categoric	2,060	1.6058	0.4888	1	2
m5b_q03	Work is physically demanding, categoric	2,057	4.8799	2.4205	1	10
Occupation	Occupation ISCO Rev 8, categoric	2,063	5.1745	2.0863	0	9
	Measures of schooling (investment in skills)					
age_start	Age at first grade, continuous	1,979	6.7019	1.0377	3	18
years_educ~t	Number of years of schooling, continuous	2,053	10.7964	4.4369	0	22
	Measures of background characteristics					
father_e~456	Indic of father's post-sec schooling, dummy	2,063	0.1924	0.3943	0	1
ses_1	Ind of low ses at age 15, dummy	2,057	0.2426	0.4288	0	1
	Controls					
Gender	Indicator of female gender, dummy	2,063	0.4392	0.4964	0	1
Age	Age in years, continuous	2,063	31.8391	9.7241	15	64
Bmi	Body mass index, continuous	1,984	24.5175	4.18401	14.7	56.1
	Instruments: p1985_ X Quarter of Birth					
qob_2_	Interaction of reform dummy and Q2	1,995	0.2942	0.4558	0	1
qob_3_	Interaction of reform dummy and Q3	1,995	0.2541	0.4355	0	1
qob_4_	Interaction of reform dummy and Q4.	1,995	0.1975	0.3982	0	1

Note: Table reports summary statistics of all variables for the employed in urban Kenya. Source: Author's elaboration World Bank's STEP data for Kenya.

Measures of Labour Market Outcomes:

For earnings, please see the Data Subsection of Chapter 3. I have reported the summary statistics for hourly earnings in USD but in all analysis in this chapter I have used log hourly earnings in USD. Earnings (log hourly) is the sole variable of the first dimension of the labour market outcomes of interest. The second dimension is the formal employment opportunity: this is measured using three variables, firm size or number of employees, which ranges from a single employee (1) to over 200 employees (7); affiliations with social security in employment, yes (1) and no (2); signing of contract, yes (1), a written agreement/appointment letter (2), and no (3). Except for firm size, I have re-coded other variables for consistency. Hence, for affiliations to social security in employment, I recoded this to (1) for no and (2) for yes. For employment which involves signing a contract, I took the reverse, re-coding with (1) for no and (3) for yes. A major rationale for including this dimension is an attempt to fully understand the extent each of the predictors explains the formality[116] of employment in a region where over 75% of the employed are informally employed. The effect of having an informal or 'black' economy adversely impacts growth and development. Hence it would be of use and policy relevance to understand drivers of informality in employment.

The final dimension of labour market outcomes is the safe working environment, this is measured using three variables: the first is a measure that indicates if a worker had lifted anything weighing over 50lb at work, with yes (1) and 2 (no). The second variable is an indicator of the perception of how physically demanding a job is with 1 indicating the least and 10 for most. For consistency, this is recoded to (1) for most physically demanding to (10) least physically demanding. The third variable in this dimension is the occupations of respondents, ranging from most unsafe (military) to least unsafe (those in managerial positions). This is an attempt to fully understand how each of the predictors in this study explains being employed in a safe environment. The physical intensity of a job is crucial for health-related risk factors. Although this is more in the health-related literature (please see Widanarko et al., 2015a and Widanarko et al., 2015b) such factors are increasingly important considerations in assessing job remuneration and satisfaction. Understanding drivers of a safe work environment in this sense is of useful policy relevance, especially in regions of the world where humans would normally do jobs that are usually done by machines, due to weak technology. Again, understanding the drivers of a Safe Work Environment will support employment and labour conditions of service which can impact health outcomes.

[116] See page 136, Table 3.3 for the final edit for this descriptive evidence of the formally/informally employed in urban Kenya.

Measures of Investments in Skills (Schooling) and Related Instruments:

I take t=1 at the pre-school level, this is defined by the variable, age at first grade (expected age of school commencement), hence, (age_start); and t=2 at the entire formal schooling interval, hence, the actual number of years of schooling. As instruments of schooling or investments in skills, whilst the study of Krishnakumar and Nogales (2020) inspired by Cunha et al. (2010); Cunha and Heckman (2007; 2008); Trostel et al. (2002) have considered background characteristics such as socioeconomic status, number of economic shocks and number of siblings at age 15. However, to assess exogenous variation in schooling and be consistent with estimates in analysis in reduced forms I instrument schooling with variables deemed to result in 'exogenous' variation in schooling. I use the three interactions of each of the quarters of birth and the reform dummy (p1985_) as instruments (see Chapter 2). Please see further descriptions and descriptive evidence of all variables in the Data Sections of Chapter 2 and the Appendix Chapter.

The Controls

As controls, I account for age, gender, and BMI, I only control for the latter and do not attach meaning to estimate due to the specification of the variable.

Background Characteristics

Background characteristics for the structural relations in System 10 (see model specifications) include, $p, make\ measures\ of$ parental skill; $and\ \bar{S}_3$ which are measures that impact earnings and employment. I explore parental (father) educational attainment as a measure of parental skills. I create dummies as indicators of parents with post-secondary education, a further background characteristic of interest in the skill structural relation is the socioeconomic status at age 15. I do these for consistency with the reduced form analysis (see Chapter 2).

5.3 Results and Discussions

As earlier highlighted, in the previous analyses (chapters 2 – 4), several limitations in using a single cross-section of data; and the methods (even after correcting for endogeneity issues) inhibit making firm conclusions on the findings.

5.3.1 Results

Table 5.2 shows outputs from estimates of the structural relations underpinned by the outputs of the measurement relations in Table 5.3.

Table 5-2 Estimates of Structural Relations (Standardised)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Outcome Variables (Full Sample)						
	Employment			Skills		Schooling	
	Earnings (f3)	Formal Emp Opp (f4)	Safe Work Environ (f5)	Non-Cognitive (f2)	Cognitive (f1)	Educ. Attain	Pre-Sch Years
Non-Cognitive Skill	0.602*	0.686*	0.579*				
	(1.961)	(1.683)	(1.094)				
Cognitive Skill	0.059*	0.096**	0.124**				
	(0.031)	(0.031)	(0.026)				
Pre-School Age (Age in Grade 1)				0.026	-0.130***	-0.157***	
				(0.004)	(0.021)	(0.076)	
Educ. Attain (Year of Education)				0.731*	0.448***		
				(0.014)	(0.009)		
Father's Post-Sec Sch				0.181*	0.086*	0.354***	-0.080
				(0.022)	(0.072)	(0.340)	(0.104)
Low SES at age 15				-0.042	-0.008	-0.149***	-0.001
				(0.010)	(0.056)	(0.237)	(0.086)
Age	0.197***	0.145***	0.093**				
	(0.003)	(0.003)	(0.003)				
Gender	-0.070*	-0.073*	0.067				
	(0.062)	(0.055)	(0.041)				
BMI	0.022	0.031	0.007				
	(0.007)	(0.007)	(0.005)				
P1985#Q2						0.067	0.072
						(0.402)	(0.112)
P1985#Q3						0.072	0.100
						(0.418)	(0.128)
P1985#Q4						0.073	0.154***
						(0.447)	(0.119)
Constant	0.000	0.000	0.000	0.000	0.000	3.939***	5.920***
						(0.844)	(0.258)
R2	0.442	0.545	0.415	0.681	0.279	0.205	0.026
Observations (1108)							

Note: Table reports estimates of the structural relations. Columns (1)-(3) present the output of the three dimensions of labour market outcomes as (f3) for earnings outcome as log hourly earnings in USD; (f4) for formal employment opportunity which is measured with firm size, affiliations with social security and employment that involve the signing of a contract or a formal agreement of a sort; (f5) for safe work environment is measured using perception of safety on the job, a measure that shows if respondents lift over 50lb weight at any time on the job and lastly how safe the occupation is. Columns (4)-(5) present outputs of the dimensions of skills outcome as (f2) for non-cognitive skills measured with Big 5 Personality Traits; and (f1) for cognitive skills measured with the ten (10) PVs of reading proficiency. Whilst (f1) - (f5) are the outputs with latent outcomes, columns (6)-(7) are the outcomes of schooling or investment in skill. (6) is the output with outcome as educational attainment as years of schooling and (7) is the outcome of pre-schooling. The predictors are as defined in System 10. p-values in parentheses: * p<0.05; ** p<0.01; *** p<0.001.

Table 5-3 Estimates of Measurement Relations (Standardised)

Indicators	**Associated Latent Variable**				
	Col (1)	Col (2)	Col (3)	Col (4)	
	(f5)	(f4)	(f2)	(f1)	R2
$\boldsymbol{Z_3^C}$					
pvlit1				0.965***	0.932
pvlit2				0.978***	0.957
pvlit3				0.964***	0.930
pvlit4				0.969***	0.939
pvlit5				0.968***	0.937
pvlit6				0.978***	0.956
pvlit7				0.977***	0.955
pvlit8				0.976***	0.952
pvlit9				0.977***	0.954
pvlit10				0.971***	0.946
$\boldsymbol{Z_3^{NC}}$					
agreableness_av			0.127		0.016
conscientiousness_avg			0.148*		0.022
openness_av			0.415**		0.172
stability_av			0.067		0.004
extraversion_av			0.154*		0.024
$\boldsymbol{Z_3^{Y4}}$					
firm_size (m4c_q20)		0.518***			0.544
social securit (m4c_q21)		0.953***			0.909
signed contra (m4c_q18)		0.856***			0.732
$\boldsymbol{Z_3^{Y5}}$					
physical demand at work	0.561***				0.314
heavy lifts at work	0.673***				0.453
risk_occupation	0.904***				0.817
number of observations			(1108)		

Note: The table reports the estimates of the measurement relations. Columns (1)-(4) present the outputs of four (4) measurement relations as the respective outcome variables in Table 5.2. With (f2) for formal employment opportunity which is measured with firm size, affiliations with social security and employment that involves the signing of the contract or a formal agreement of a sort; (f3) for the safe work environment is measured using the perception of safety on the job, a measure that shows if respondents lift over 50lb weight at any time on the job and lastly how safe the occupation is. The predictors are as defined in System 10. p-values in parentheses: * p<0.05; ** p<0.01; *** p<0.001.

5.3.2 Discussions

Effects of Background Characteristics on Schooling and Skill

Assessing the effects of background characteristics (proxied by father's education and socioeconomic status) on schooling and skill in line with the research questions of Chapter 2. Consistent with the argument raised on the effects of background characteristics on the skill, the findings from Table 5.2 suggest background characteristics, particularly, the indicator of the father's post-secondary schooling raise skills (both cognitive and non-cognitive) directly, this suggests that the impact of the father's post-secondary education on skill is through its effects on schooling, specifically, see columns (4)-(7). The previous findings, particularly, those from the Difference-in-Differences and Instrumental Variables analyses of Chapter 2 give some insights into this. However, this is the first useful evidence that shows (directly) a positive effect of a father's post-secondary education on adult skills. This is consistent with the literature on intergenerational transmission/mobility of schooling and abilities. The effects of background characteristics on educational attainment and age-at-first grade (pre-school-age) are as expected. However, the effect of low socioeconomic status on skill is statistically insignificant, amidst the statistically significant (and adverse) effects on schooling. This is consistent with the outcome of the Difference-in-Differences, see Chapter 2.

Pecuniary returns to skills

Examining the effects of skills (cognitive and non-cognitive) on earnings and other dimensions such as safe work environment; and formal employment opportunity. Interestingly, the evidence suggests, although less precisely estimated (with high standard errors relative to cognitive skills), non-cognitive skills impact formal employment opportunity, earnings and safe work environment, in this order. Cognitive skills favourably impact a safe work environment, employment opportunity, and earnings. Having evidence of positive effects of non-cognitive skills across all dimensions of labour market outcomes of interest (in this study) accentuates the growing importance of personality traits for employment. This is well documented in the OECD settings. Similarly, in Kenya, cognitive skills are crucial for all dimensions of the labour market outcome of interest in this study, particularly, for safe work environment and formal employment opportunities, which accentuate its importance for jobs in large organisations with favourable work conditions. Interestingly, in Bolivia, the study of Krishnakumar and Nogales (2020) finds that cognitive skills impact all dimensions of their measure of labour market outcome positively, suggesting the importance of cognitive skills over non-cognitive skills for 'work-related well-being' that make the labour market outcome of interest in their study.

Both cognitive and non-cognitive skills have statistically significant effects on earnings. Comparing the effects of cognitive and non-cognitive skills in this analysis to the previous analysis. For the non-cognitive skill, the evidence suggests findings from the structural and measurement relations are consistent with OLS evidence of Chapter 3 that argues the OLS evidence is causally identified, particularly for the effects of openness_av, extraversion_av and conscientiousness_avg (see measurement relations in Table 5.3). For cognitive skill, whilst the findings from the structural relations show some consistency from our previous analysis in Chapter 3 where the Instrumental Variables approach, at best, only shows a positive effect on the return to skill of the male gender. These findings—the (overall) weak effects of cognitive skills from the reduced form analysis (IV) and the structural relations—unravel useful insights and may strongly suggest cognitive skills are either less rewarded or not considered, amidst their scarcity in the labour markets of developing contexts. This is consistent with the argument of Pritchett (2001) that asserted cognitive skills or skills from schooling may have some private returns through socially unproductive activities as it is seldom applied in productive activities in developing contexts, amidst its scarcity from the school system (see Pritchett (2001)). Pritchett (2001) went on to cite several cases, especially in public services where cognitive skill from education is misplaced or not put to good use, using these to support the claim that this is at least in part, the reason for the poor economic growth and development of the region.

Effects of Education on Skills

The evidence from the structural relations shows substantial effects of schooling on cognitive skills (statistically significant at the 0.1% level) and non-cognitive skills (statistically significant at the 5% level) in urban Kenya. Both effects are precisely estimated with relatively low standard errors. This result is consistent with the effects of schooling on cognitive skills from Chapter 2. As expected, at the preschool level, the effects on cognitive skills are negative and statistically significant at 0.1%, further suggesting schooling (substantially) results in cognitive skills, whereas, at the preschool level, although not negative, the effect of non-cognitive skills is not different from nil (hence, not statistically significant). Overall, at the post-school attainment levels, the evidence suggests, that schooling results in cognitive and non-cognitive skills in urban Kenya. Whilst these estimates of the effects of cognitive and non-cognitive skills in Kenya are consistent with those of Bolivia (see schooling Krishnakumar and Nogales, 2020), some evidence from Kenya slightly differs from those of Bolivia.

In Bolivia the effects of schooling on cognitive and non-cognitive skills are statistically significant relative to non-cognitive skills are statistically significant at the 0.1% level. However, in Kenya, the effects of years of schooling on cognitive skills is statistically significant at the 0.1% level and the effect on non-cognitive skills is at the 5% level. Based on the model specifications (see analytical framework), the effects of schooling on these skills define the mechanisms through which schooling or investment in schooling impacts labour market outcome. The evidence suggests the substantial effects of schooling on cognitive skills and non-cognitive skill is reflected across the dimensions of the labour market outcome in Bolivia as in Krishnakumar and Nogales (2020). However, this is not exactly the case in Kenya, where skills (particularly, cognitive skills) appear misplaced as Pritchett (2001) suggested. From the structural relations (Table 5.2), relative to the statistically significant effects of schooling on cognitive and non-cognitive skills; the effects of these skills on earnings (as a dimension of the labour market of interest) are inconsistent, this is the case of cognitive skills where effects on earnings are substantially less statistically significant.

5.4 Summary and Concluding Remarks

Demystifying the link across education, skills and labour market outcomes in the non-OECDs unravels useful insights, however, this presents some econometric challenges owing to the data limitations that plague useful research in the non-OECDs. Following a suitably adapted dynamic framework, pioneered by Cunha and Heckman (2007)—the Technology of Skills Formation— the study of Krishnakumar and Nogales (2020) in Bolivia set the scene using an adapted version of this dynamic framework, presenting useful insights for overcoming such data limitation faced in this study. I deploy a similar approach to unravel useful insights into the relationships across schooling, skill, and labour market outcomes in Kenya, amidst the data limitation challenge presented using the available single cross-section of data used in the analysis.

In estimating the effects of the father's education on the schooling and skills of respondents, whilst accounting for the latency of skills, in a dynamic framework. The key findings affirm the argument in support of the literature on intergenerational education and skill mobility, as the effects of background characteristics—particularly, the father's post-secondary schooling—impact the skills of their offspring by first impacting their schooling, in urban Kenya. This gives (further) support to the argument for 'equity in access', over 'quality in' schooling, to effectively raise skills in urban Kenya.

Estimating simultaneously in a dynamic framework, the effects of schooling from skill; and in turn, the effects of skills on labour market outcomes. Taking skills as the mechanism through which schooling impacts labour market outcomes. Key findings are consistent with findings that individually estimate the (aforementioned) relations in reduced forms. Whilst the evidence suggests schooling substantially impacts skills, the effects of skills on earnings although statistically significant at a 5% level. This is generally inconsistent with the effects of schooling on skills. Suggesting that the effects of schooling on earnings are at best, in part, what explains earnings, in urban Kenya. This suggests skill is misplaced in the Kenyan labour market, as argued (by Pritchett, 2001), particularly in sub-Saharan Africa. This is unlike the Bolivian labour market where the effects of schooling on cognitive and non-cognitive skills are consistent with the effects of cognitive and non-cognitive skills on earnings (see Krishnakumar and Nogales (2020).

6 Conclusion to the Thesis

One of the main novelties of this study is the consideration of ‘skill’ as an additional dimension of human capital, as opposed to mere schooling. The consideration for skill dimension is deemed necessary owing to the context of this study. The Big 5 are the main measures of non-cognitive skills and reading proficiency (in Plausible Values) is the sole measure of cognitive ‘skill’. The analytical chapters of this thesis have addressed interrelated research questions on returns to education and skills for Kenya. The study examines the education-skill-earning-growth link as discussed in Chapter One—the introduction to the thesis. This involves thoroughly investigating the following relationships: education-skill; education-earnings; and skill-earnings and the related externalities of schooling and skills. Most of these relationships are examined in reduced form, and in a dynamic framework. The latter tests the robustness of outcomes in reduced form and improves the internal and external validity of estimates. Hence, an attempt to overcome the limitations of a single cross-section of data, in this study. The findings of this study have useful implications for policy and future research.

The contributions—policy and the literature—of this study are far-reaching. I will now highlight the contributions, policy relevance and implications for future studies. The main contributions of this study include providing useful estimates of private returns to the measures of human capital—education and skill—and the related externalities (effects of aggregate schooling skill) that impact estimates of social returns. The main outcomes are cognitive skills (proxied by reading proficiency) and earnings. Useful estimates of the returns to education and skill can aid policymaking on several fronts, particularly, the decision or approach to investing in education by the government, donors, firms, and individuals and this can impact the related policymaking that relates to employment in labour markets and for education across districts. Also examined is the impact of education reforms, and family background characteristics on educational attainment and skill acquisition. This is argued to have specific policy implications, which relate to access to (and quality of) education, particularly, raising awareness and understanding of the efficiency of schooling—the extent to which schooling explains skill—and human capital development in a developing context. Next, providing new estimates of returns to cognitive, and noncognitive skills in urban Kenya adds to the literature and supports useful comparisons with existing estimates in the OECDs. This supports an assessment of human capital from education, and the labour market, in urban Kenya. As earlier mentioned, the links across education, skill, and productivity, cannot be over-emphasised as governments can use related reforms or policies to develop and harness human capital and ultimately, achieve useful economic growth and development. Besides the implications for policymaking, findings from this study have implications for future research. In the following subsections, I will summarise the main findings in the chapters of this study. I then turn to discuss key findings particularly those with interconnected themes across testable predictions of the chapters of this study. Then I discuss limitations, policy implications and implications for future research.

Main Findings, Limitations and Implications for Policy and Future Research—Individual Chapters

In the first analytical chapter, the findings suggest that having a father without post-secondary schooling shows evidence of upward mobility in the schooling of wards (respondents), a father's post-secondary schooling has no statistically significant effect on the skills of their ward. However, further evidence suggests that this is partly due to the weak quality or inefficiency in education, in urban Kenya—where those with minimal schooling have higher skills, relative to those with high educational attainment. Further evidence suggests that the substantial rise in the schooling and skills of those who have fathers without post-secondary education (the disadvantaged) is attributable to the reform. Hence, it is reasonable to conclude that it is of policy relevance that effort to raise aggregate education (average schooling) is crucial for skill than efforts to raise quality in schooling. Furthermore, useful consideration of background characteristics that should support interventions that raise skill through increased access to schooling is the abnormal and low-skill profiles in developing contexts. Besides the findings on the effects of background characteristics on schooling and skill and the effect of schooling on the skill of the disadvantaged, as earlier discussed, comparing the non-parametric evidence (see Figures 2.5 and 2.6) and those of Figure 2.3 suggests the need to further raise schooling (especially for the disadvantaged) for useful skill. The evidence from the non-parametric analysis suggests that besides background characteristics, other unobserved factors are responsible for the abnormal skill profile in Kenya. Besides the non-parametric evidence, the parametric evidence—the Oaxaca-Blinder Decomposition, the 2SLS-IV estimates — together with an opportunity for higher quality (such as private) schooling that shows evidence of the sharp differences in schooling and skills across urban Kenya. Whilst the reform is very much directed to the disadvantaged and impacting their schooling and skill (as seen in the 2SLS-IV estimates, see Table 2.20), the Oaxaca-Blinder Decomposition shows sharp differences in schooling between the advantaged and disadvantaged is a crucial driver of the differences in skills between these groups. Further evidence of the accessibility of quality or private schooling that suggests the advantaged have better access to such (private) schooling that explains more years of schooling for the advantaged relative to the disadvantaged is another evidence that can reasonably explain the substantial differences in skill between the advantaged and the disadvantaged. Please see Tables A1.7.1 and A1.7.2 in the appendices. Overall evidence shows that for more useful skill distribution, a plan for increased schooling taking account of the background characteristics of respondents, is crucial for skills, in Kenya—this is evident across the non-parametric and non-parametric evidence of this study. However, this study is limited in this regard as the abnormal (double- or multi-peaked) skill profile is at best, partly explained (see, Fig 2.5 and 2.6; compared to Fig 2.3) by differences in accessibility to schooling; inefficiency in schooling and the effects of the reform across on individuals based on their background characteristics. Hence, as an implication to future research, further studies that will investigate other factors that could fully explain, the abnormal skill profile in Kenya and sub-Sahara Africa at large would be useful to strengthen policymaking.

Another crucial point of policy relevance, with some implications for further studies is accounting for the effects of internal migration. The evidence (please, see Table A1.10 of Appendix A1) suggests that, relative to respondents who studied and are working in the same district, moving to an urban district (possibly, either moving from rural or from smaller urban districts) after studies is associated with a shortfall of between 0.8-1.2 years of schooling which in turn explains a standard fall of between (14-32) % in reading proficiency. These effects of internal migration are statistically significant at the 0.1% level. A policy implication of such loss in schooling and skills attributable to internal migration includes the need to raise schooling across districts to maintain the supply (balanced with demand) of quality of labour across districts. Hence, this entails an approach that will maintain a useful supply of labour, by putting in place structures to keep useful demand whilst maintaining a useful quality of the workforce. If the cities where migrants get jobs are not in shortage of skills or do not have excess demand over supply of workforce, the migrants will not (easily) find jobs; alternatively, if the migrants can find gainful employment where they studied, they will do so and not have to emigrate. Hence prioritising an even distribution of schooling (hence, skill) such that all districts attain similar or same average schooling will not only aid a more even development or growth across districts, but it will also maintain useful quality of labour across the country and give opportunity for more growth as employers can easily find workers with the right education/skill. Lack of data impacted this analysis as it would have been better to use additional or more useful measures to capture migration but limitations in the dataset (no useful variables), meant this was not very possible.

In the second analytical chapter, key findings suggest that the OLS (Ordinary Least Squares) estimates of non-cognitive skills are substantially robust to controlling for schooling. Estimates are consistent, with Openness to Experience and Conscientiousness yielding positive and statistically significant wage effects on which causal inferences are drawn. This is inspired by the study of Mosca and Wright (2018) who showed that these traits are seldom endogenous in an Ordinary Least Square Estimation. The evidence suggests that Openness to Experience has the strongest effect, with a standard 35.9% rise in hourly earnings, statistically significant at the 0.1% level, whereas a standard rise in Conscientiousness has a wage effect of 12.6% rise in hourly earnings, statistically significant at the 5% level. However, this study admitted the latency of skill without which estimates of the wage effects of cognitive and non-cognitive skills would be spurious because the data is collected when the respondents became adults. The robustness of the OLS estimates was re-examined in a dynamic framework (chapter 5) that accounts for the latency of skill, and the OLS estimates are shown to be sufficiently robust. Furthermore, the 2SLS-IV estimates show some robustness (consistency) in the private return to schooling and cognitive skills. These initial findings suggest, for the main analytical sample (pool), no statistically significant wage effects for schooling and skill from which causal inference can be drawn. Further evidence from subsampling (heterogeneity analysis) shows that relative to the female gender, the male gender has some useful returns to their schooling and cognitive skills, with an additional year of schooling explaining a 25.6% hourly wage rise, statistically significant at the 1% level. For the measure of non-cognitive skill, reading proficiency (unstandardised), the evidence suggests, a unit rise in the PV (plausible values) explains a 0.77% hourly wage rise, statistically significant at the 5% level. Using the baseline estimates, the Oaxaca-Blinder decomposition suggests a 23% wage differential across genders. Differences in schooling and skill (human capital characteristics/endowments) between genders explain about thirty-seven per cent of the wage difference between males and females. A substantial proportion of the wage differential between genders is due to potential discrimination. No evidence of (potential) discrimination is attributable to cognitive skills or schooling, however further evidence suggests that the potential discrimination in wages between genders comes through their non-cognitive skills (or personality traits), particularly, women are rewarded over men for their Consciousness, and men are rewarded over women for their Openness to Experience. Some limitations of this study stem from using hourly earnings, across all categories of the employed[117]. These categories of the employed have very different

[117] The categories of the employed include the following: wage—public and private wage-employed; self—lone-employed and entrepreneurs; formality—the formally- and informally- employed).

compliance with taxation, and as expected, there is high variability in earnings attributable to differences in taxation as these earnings are calculated from the self-response survey instruments. This can introduce substantial bias in the estimates of returns, across the categories of the employed. However, careful consideration is given in interpreting outputs and comparisons are much more considered across similar categories based on their possible tax status. By this, I mean take time to compare the wage-employed (private-public wage employed); and the self-employed (lone-employed and entrepreneurship) avoiding comparing findings across the wage- and self-employed.

In the third analytical chapter, initial findings suggest that, for the pecuniary external returns to schooling and skill, the 2SLS-IV output suggests the effects of aggregate schooling and skill on earnings are not different from nil, hence, statistically insignificant. However, substantial heterogeneity is evidenced, specifically, a one-year rise in average schooling of males (and the private-sector wage employed) across districts in urban Kenya shows a 25.3% (and 23.2%) loss in their hourly earnings, statistically significant at the 5% level. No statistically significant effect on the pecuniary externalities of skill exists (based on the 2SLS-IV). Inspired by Liu (2007) who suggests the sum of external and private returns should give an approximation of pecuniary social returns to education—the effects of average schooling on an individual’s earnings. Using the same 2SLS-IV specifications, with an external return to an additional year of increase in aggregate schooling for males and the employed of -25.3% (and -23.2%) as earlier stated. With a private return of 18.2% (and 20.1%) respectively. This results in an approximated social return of about -7.1% (-3.1%) for the males and employed respectively. The negative social returns suggest that, purely on financial considerations, the government of Kenya is currently overinvesting in schooling. Interestingly, based on the main analytical sample (the pool), the non-pecuniary human capital externality suggests that a one-year increase in aggregate schooling in urban Kenya results in a standard loss of 24.5 percentage points in reading proficiency. With a rise of 31.8 percentage points in cognitive skill for an additional year of schooling for an individual, as an extension to pecuniary social returns to schooling to human capital, an approximation of non-pecuniary social returns to education—the effects of aggregate schooling on aggregate skill—suggests that a positive social return of education on skill, further suggesting the need for more investment in schooling for skills in urban Kenya. However, after a re-examination of the tests of the instruments (see Appendix A3) in the 2SLS-IV estimates, the evidence suggests the outcomes of the 2SLS-IV estimates may have been spurious. The OLS outcomes showed useful robustness, particularly in capturing pecuniary and non-pecuniary externalities of schooling with quadratic terms this accentuated some useful conceptual underpinnings of this study—particularly, the argument of more schooling for skill. Turning to the OLS outcomes, the evidence suggests, consistency in the pecuniary and non-pecuniary external and private returns with substantial evidence of skills from schooling (particularly for the disadvantaged), which further explains earnings. The negative externality becomes less negative (or positive), with a rise in average schooling, across districts. Hence, individual and aggregate schooling, explain skills, and in turn, impact earnings. However, this comes with some exceptions—the public service wage employed and those in entrepreneurship—where schooling and skill yield little or no externalities. The evidence for those in entrepreneurship and those in public service wage employment suggests schooling is privately remunerative but socially unproductive as private returns to schooling are substantially high, however, the effects of the externalities of schooling are not different from nil. Furthermore, estimates of private and external returns to skill are not different from nil. This is regardless of the threshold or level of schooling attained. Besides this anomaly for the public-sector wage-employed and those in entrepreneurship, the overall findings suggest that effort to raise schooling raises skills and earnings in urban Kenya. This study adds to the literature by providing robust estimates of private, social, and external returns to schooling and skill and improving understanding of Pritchett's (2001) propositions, in the context of urban Kenya. An important policy insight offered by this study is the requirement to further invest in average schooling for skill in Kenya, particularly, in the light of the abnormal outcomes for the public service wage employed and those in entrepreneurship. Hence, further policy contributions drawn, relate to the emphasis on skill in public service wage employed and those in entrepreneurship. Finally, a need for Entrepreneurship Education and Training. These insights are crucial for policymaking on growth and development, in urban Kenya owing to the strategic importance of these employment categories. Finally, this study strongly emphasises (and advocates) the need for researchers to use both (and not either) measures of schooling, and skill to fully capture the effects of human capital, particularly, in developing contexts.

The final analytical chapter suggests that deploying the dynamic framework, the evidence from the structural relations (and the associated measurement relations) is consistent with the estimates in reduced form. However, the evidence from the Difference-in-Differences (DiD) estimations in Chapter 2 shows that the indicator of the father's post-secondary schooling impacts schooling and not the skills of the offspring. Using the dynamic framework, I affirm that having a father with post-secondary schooling does not only explain schooling but skill. Highlighting strong evidence of at least persistence (or upward mobility) in education and ability (skill) between a father and their offspring. This accentuates an intergenerational mechanism for educational attainment and skill proliferation that should not be overlooked in policymaking, relating to education, skills, and employment. This structural estimation overcomes the drawbacks in data, particularly, the issues with instrument validity, age-cohort confoundedness, and omitted variable biases raising the internal and external validity of estimates in Chapters 2 and 3 but not Chapter 4. A major limitation of this study is the narrow definition of cognitive skills. Typically, studies used several other dimensions of cognitive skills to assess the effects of cognitive skill, but this study and this chapter have defined skill strictly by reading proficiency in plausible values. Although these are known to be robust measures free from reticence, however it is not the norm to use a single measure of cognitive skills (see Krishnakumar and Nogales 2020). Krishnakumar and Nogales (2020) used several other measures, like numeracy in addition to reading proficiency. This study attempts to capitalise on tests, instead of self-response survey items acknowledge the bias that may come from reticent respondents. In the STEP data, only reading proficiency is a direct measure of cognitive skill others are self-response (or indirect) measures of cognitive skill. However, I acknowledge the data limitations here.

Further Findings, Limitations, and Implications for Policy and Future Research — Interconnected Themes (Across Chapters)

Examining the education-skill-earnings-growth link unravels several interesting findings across the interconnected themes of this study. Firstly, examining the effect of individual schooling on individual skill (chapter 2); and the effect of aggregate schooling on individual skill (chapter 4). Although this was briefly discussed as 'non-pecuniary returns' in chapter 4. However, to make some concluding remarks on this, I will now emphasise some salient points here. Whilst findings from this study are in concordance with Pritchett's (2001) propositions that suggest schooling is privately rewarding but socially unproductive, this analysis critiques the findings of several related studies, including the propositions of Pritchett (2001) that suggest schooling yields little or no skill in most developing contexts. Interestingly, whilst the baseline (OLS) evidence for urban Kenya suggests individual schooling yields substantial skill, there is an adverse effect of aggregate schooling on individual skill, this effect is strongly dependent on the level (number of years) of average schooling[118]. The evidence suggests not keeping average schooling at a certain threshold can

[118] the average or aggregate schooling across the districts appear to be a major driver of the overall (regardless of categories of schooling) and adverse effects of schooling beyond lower secondary schooling. Interestingly, average schooling enters as quadratic terms (U-shaped). Hence, the adverse effect of average schooling become less adverse as average schooling rises, hence, this effects of average schooling on skill may become non-

mar the effects of schooling on individual skills. Specifically, column (6) of Table 2.13 suggests an additional year of schooling for an individual explains a 13% standard rise in reading proficiency. However, an additional year of aggregate schooling explains a 25.6% standard fall in reading proficiency, this assumes a nil level of average schooling. Specifically, the evidence suggests if average schooling is not ≈12 years of schooling, the effects of individual schooling on their skill will be offset by a negative effect of the externality (effect of average schooling on individual skill) to give an estimate of the social returns to skill. From the findings (as above), the adverse effect of aggregate schooling on individual skill may be much weightier than the favourable effect of individual schooling on individual skill, think about a district as Trans Mara (see Table 4.1) where average schooling is only 4.8 years of schooling, this means, all things being equal, average schooling would have to be raised by over seven (7) years of schooling for individual schooling to have its full effect on skill. Suggesting that, in some cases, the combined effects of average and individual schooling on individual skill may have an adverse effect. Hence, this suggests that schooling (individual, and in aggregate) can adversely impact reading proficiency in urban Kenya[119]. However, the evidence from the analysis on the 'efficiency of schooling' suggests that additional schooling beyond the ISCED2 level of schooling first results in a fall and then a rising reading proficiency follows with more schooling. In addition to this, being consistent with the quadratic nature of average schooling (U-shaped), it does not suggest that schooling should terminate at ISCED2. This suggests a policy conclusion, that further investment in skill (through access to schooling) is worthwhile, this should always be the case for urban Kenya. The combined effects of individual and average schooling on individual skill are considered as the first approximation of 'non-pecuniary social returns' argued to be a useful basis for further investment in schooling (see Chapter 4). The quadratic nature of average schooling suggests if this approximation of social returns is negative, the usual understanding of overinvestment in schooling does not hold, rather, further investment to increase average schooling would be worthwhile, regardless of private returns or externalities of schooling.

---

negative at a certain threshold of average schooling. This presents a very interesting outcome in support of some predictions of this study, particularly the argument for increased access to schooling for skill, regardless of the quality of schooling. Suggesting a substantial increase in access or average schooling will result in a positive effect of schooling on skill on the long run.

[119] this is consistent with the findings on efficiency of schooling, see columns (7)-(12) of Table 2.13 where relative to low levels of schooling, an additional year of schooling at high levels of education explain low reading proficiency.

The baseline (OLS) estimate is the choice specification for the externality of schooling on skill as the 2SLS-IV estimates are shown to be defective due to the weak instruments used. Table 2.20 presents the 2SLS-IV estimates, column (1) suggests that the overall (individual, and aggregate) effects of schooling, on skill are positive as the reform-affected estimates show that an additional year of individual schooling explains a standard rise of 31.8% in reading proficiency. However, an additional year of average schooling only explains a standard loss of 24.5% in reading proficiency. The overall (first approximation of social returns) is positive. However, accounting for parental education (by taking subsamples) in columns (3) and (5), the evidence suggests the overall reform-affected returns to schooling for the disadvantaged are consistent with baseline (OLS) outcome where the adverse effect of aggregate schooling is greater than the favourable effects of individual schooling. Overall, this results in an adverse effect of schooling on skill. As earlier argued the baseline (OLS) outcome is considered robust, relative to the 2SLS-IV specification. Firstly, this is partly due to its conceptual underpinning—the quadratic effects strongly reveal evidence of the need for more (access to) schooling for skill in sub-Sahara Africa—hence, the robustness of the OLS specification over the 2SLS-IV specification, as the latter fail to account for the quadratic form of the average schooling. Secondly, the overall weakness of the instruments used for instrumenting individual and average schooling in the 2SLS-IV specification suggests that the outcome of the 2SLS-IV cannot be fully relied upon due to this substantial limitation of the 2SLS-IV approach. Please, see Tables A3.1-A3.7 of Appendix A3 for more on this, showing at best, the outcome of individual schooling is supported in the 2SLS-IV as average schooling is better taken to be exogenous. Furthermore, crucial for policymaking for increased skill from schooling is to widen access or raise schooling as much as possible. Particularly, in the simple calculations based on the OLS estimates for average schooling, taking the first-order partial derivative of the quadratic outcome of average schooling (see column (6) of Table 2.13) suggests the current average schooling (see Table 2.1) of about 10.3 years of schooling would have to rise to about 11.9 years of schooling to bring the effect of the externality of schooling on skill to a non-negative effect. Hence, an increase in average schooling of 1.6 years, across the districts of urban Kenya, is required to keep a non-negative externality of schooling. The inefficiency in schooling may warrant increased quality inputs, evidence from the nature and effect of the externality of schooling suggests a further increase in average schooling further raise skill suggesting the idea of raising access that ultimately raises average schooling is a viable strategy for increased skill as earlier argued. Hence, prioritising increased access over increased quality is of dire policy relevance for useful skill proliferation. Hence, raising years of schooling over teacher training would be ideal, especially in situations of scarce or limited resources. Beyond this policy implication, an implication of some of these to future research is the need to obtain a more robust dataset that can result in useful exogenous variation in aggregate schooling (not just individual schooling) that can raise or support an assessment of the internal validity of the reform-affected estimates of average schooling in the 2SLS-IV specification, as doing so will further ascertain the robustness of the related OLS estimates. This is an area to be further explored in subsequent research as the structural equation model (the dynamic framework) used to assess the effects of individual schooling on skill did not account for aggregate schooling.

Other important interconnected themes across the analytical chapters are the effect of individual schooling on individual skill (as in Chapter 2); and the effect of individual schooling on earnings (as in Chapter 3). Chevalier et al., (2004) suggest that the difference between the wage effects of schooling; and the effects of schooling on skill give insights to the proportion of the wage effect of schooling that is 'signalling'. Similarly, to affirm support for the human capital theory or show possible effects of signalling theory in urban Kenya. Table 3.7 suggests that the employed with lower-secondary schooling earn more than those with secondary and some post-secondary schooling. With the former having mean hourly earnings of $3.11 and an average skill of 191.3PV; and the latter earning only $2.88 per hour and an average skill of 198.2PV, this descriptive evidence suggests that the mean schooling and skill of those with ISCED2 are confounding relative to those with ISCED34A. However, turning to the baseline (OLS) return estimates. Table 2.13 columns (7)-(12) suggest that an additional year of schooling for those with lower secondary schooling also explains higher reading proficiency relative to those with secondary and some post-secondary schooling. Specifically, Table 2.13 columns (7)-(12) suggests an additional year of lower-secondary schooling explains a standard marginal rise of (12.4-14) % in reading proficiency whereas an additional of secondary and some post-secondary schooling explains a standard marginal rise of (11.3-12) % in reading proficiency. In addition to this, Table 3.8 suggests that having the ISCED2 qualification (with an average of 10.2 years of schooling) explains a 44.3% rise in hourly earnings, whereas having the ISCED34A (with an average of 12.4 years of schooling) qualification explains 42.2% rise in hourly earnings. These regression estimates as opposed to the mere descriptive evidence suggest having the ISCED2 qualification not only significantly explains earnings but skill. At first, further evidence suggests this may be attributable to the vocations rather than the academic qualification (see tenure of Table 3.7) as those with lower secondary education are more experienced, in addition, the nature of their vocations (which require such experience) may be useful drivers of their reading proficiency. Interestingly, the inefficiency in schooling may also suggest, this is linked to academic qualification. However, regardless of academic or vocational tracks, this evidence of higher skill for higher earnings suggests ‘skill’ (better drive productivity, hence earnings) as opposed to mere schooling is better rewarded in the labour market in Kenya, giving support for the human capital theory, as opposed to the signalling theory. This has implications for policymaking, particularly, if issues of efficiency in schooling contribute to such findings in support of signalling or human capital theories. This should require urgent attention. It is value-destroying if an additional year of schooling explains lower reading proficiency, suggesting an effort to raise access to schooling (as discussed in the previous paragraph) may be futile as a higher level of schooling should explain higher skill levels. Addressing such inefficiency may warrant an investment in school inputs, particularly, for teacher training. See the introduction where evidence suggests whilst the teaching workforce in both primary and secondary levels of schooling are considerably well trained, this is not the case of TVET colleges (that is marred by temporary and untrained staffing) where most of those with the ISCED34A qualifications may have had their training. There is every need to raise the capacity required to sufficiently deal with the issues of inefficiency in schooling in urban Kenya. However, raising school inputs (or quality) may not take preference over increased access, especially with resource constraints. Increasing access over quality is vital for equity in access as those disadvantaged by background are reasonably supported, besides this, careful observation of the inefficiency in schooling for Kenya suggests that, although reading proficiency from schooling starts falling after ISCED2, however, it starts to rise again after ISCED34A which captures the quadratic nature of externality of schooling and strongly suggests additional schooling (with a useful or reasonable quality in place) will raise substantial skills. Therefore, an attempt to raise quality inputs should not inhibit efforts to raise access to schooling. This is of crucial policy relevance for Kenya.

Like the non-pecuniary return to schooling, the pecuniary return to schooling is similar. However, how does the pecuniary return to schooling compare to the pecuniary returns to skill? The evidence suggests that individual schooling has consistent effects (linear and positive) on skill and earnings; similarly, average schooling has consistent effects (quadratic) on skill and earnings. This evidence suggests the overall (individual and average) impact of schooling is consistent across skill and earnings. In simple terms, the impact of schooling on earnings is consistent with the effects of schooling on skills. Interestingly, this is consistent with the conclusions in the previous paragraphs. Firstly, this suggests schooling is less signalling and more consistent with the human capital theory—where skill drives productivity and reward. However, determining the extent to which this is less signalling and more consistent with the human capital theory is not the objective of this study like the study of Chevalier et al., (2004). Secondly, in the first paragraph of this subsection (on the interconnected themes of this study), the evidence suggests the combined effect of schooling, hence, the effects of 'overall' (individual and average) schooling on skill may be adverse, as the adverse effect of average schooling on skill may also be weightier than the favourable effects of individual schooling on skill. This is consistent with the overall effects of schooling on earnings as the adverse wage effects of average schooling may be weightier than the favourable wage effects of individual schooling. Again, this is consistent with Pritchett (2001) who suggests schooling is privately remunerative but socially unproductive. Table 3.23 column (14) suggest that the return to an additional year of individual schooling is a 10.7% rise in hourly earnings, however, an additional year of average schooling explains slightly under 34.4% fall in hourly earnings, this becomes less negative as average schooling across districts rises. The findings suggest a non-negative externality of schooling requires ≈10.02 years of schooling. Put together, the overall effect of schooling on skill is also consistent with the overall effects of schooling on earnings (although for a non-negative externality of schooling on skill, the evidence suggests ≈12 years of average schooling) and although privately favourable or remunerative, the substantial adverse effects of the externality of schooling means the overall effects of schooling on skill and earnings may be adverse at certain number of years of average schooling. Hence, the overall effect of schooling on skills and earnings supports growth, at high levels of average schooling. This further strengthens the argument for increased access to schooling. I now turn to the arguments on the choice of the measure of human capital.

Whilst a strand of the literature advocates years of schooling[120] (see Harmon and Walker, 2001) as a measure of human capital, for another, the emphasis[121] is on skill (see Hanushek et al., 2013). This is particularly important for studies in developing contexts, where the effect of schooling on skill is deemed highly heterogeneous across developing and developed contexts. Here, the findings suggest that the overall effect of schooling on earnings is quite distinct from the overall effects of cognitive skill (proxied by reading proficiency) on earnings. The wage effects of individual schooling and individual skills are positive and are statistically significant at the 0.1% level. At first, this shows useful consistency in the measures. However, the average effects are distinct, as the wage effects of average skill across districts are linear, positive, and statistically significant, see Table 3.23 column (10).

[120] Here, schooling is taken to mean, the time (years) spent in formal education.

[121] In recent times, the emphasis on measures of skills in lieu of other measures of educational attainment (such as the number of years of schooling spent in formal education; and mere credentials) is gaining more grounds as measures of skills are deemed more plausible measure of human capital (see Hanushek and Woessmann, 2008) that relates to economic growth, particularly, in developing contexts where schooling is argued to give little or no skill (see Pritchett, 2001).

However, as earlier discussed (see Table 3.23 column (14)), the wage effect of average schooling is quadratic and statistically significant, hence, negative and getting less negative with rising average schooling. Driven by the differences in the nature and the effects of the externalities of (wage effects of average) schooling and skill, the evidence from Kenya makes it clear that schooling and skill are distinct measures and may not be used as alternative measures of human capital. Hence, whilst the individual effects of schooling and skill are consistent, the average effects of schooling and skill are not. Therefore, the overall (average and individual) wage effect of skill appears to be favourable unlike the overall (average and individual) wage effect of schooling which may be adverse at a certain level of average schooling (also see pecuniary externalities and social returns in Chapter 4). Interestingly, the differences in earnings between the formal and the informal are driven by differences in their individual and average schooling in a manner that is consistent with the overall effects of schooling and skill as discussed. However, there are several other specific cases where heterogeneity in skill (not schooling) does not account for heterogeneity in earnings (see Oaxaca-Blinder decomposition) across subsamples of interest in this study. Particularly, whilst the heterogeneity in earnings across the male/female genders; and the entrepreneurs/lone employed are well driven by the differences in their characteristics, particularly, the differences in their schooling and skill. This is not exactly the case of the public and private sector wage-employed where the substantial differences in the earnings of the public service wage employed over the private-sector wage employed are driven by differences in the individual schooling (and not skill) of the public-sector wage-employed over the private-sector wage employed. Firstly, this strongly suggests public-sector wage employment in Kenya emphasises schooling (or qualifications/credentials) over skill. Secondly, this finding accentuates skill (as distinct from schooling), however, skill does not always account for substantial differences in earnings in developing contexts, possibly, due to all or at least one of the following: skills misplacement in the labour market; credentialization; and inefficiency in schooling as earlier discussed. A further implication of these is the relatively weak productivity of the public sector wage employed seen across developing contexts, particularly sub-Sahara Africa, where employment guarantee schemes (to incentivise schooling) by governments of these contexts can lead to dysfunctional governance and public service (see Pritchett, 2001). In support of the argument of Hanushek et al., 2013, evidence from urban Kenya suggests skill may be a more appropriate measure of human capital in developing contexts as the use of measures of schooling may be misleading. Overall, the mean effect of skill is quite small (but strongly significant) relative to the effects of schooling however the former shows more consistency with the ideas of human capital linked to productivity and growth. This has useful implications for future research, especially in developing contexts. Whilst the study of Harmon and Walker, 2001 advocates the use of years of schooling on many grounds, particularly, due to its fit with the models of human capital, most of their studies have been in developed contexts where the use of schooling or skill as measures of human capital may be consistent. The implication of the findings in this study is clear, relating more to developing contexts as the empirical evidence makes it clear that schooling does not necessarily mean 'skilling' in this context, due to the externalities of schooling and skill. Hence this does not inhibit research that makes it clear that the use measures of schooling capture the mere effects of schooling. The use of measures of schooling mustn't be expressly interpreted as a measure of human capital, without additional robustness checks with other useful and 'theoretically sound' (with useful conceptual underpinning) measures of human capital linked to productivity, such as measures of skill.

Further Limitations and Implications for Future Research

Using a single cross-section of data in this study means a higher incidence of age-cohort (age, period and cohort) confoundedness. This results in biases in estimates and mar the interpretation of outcomes. The age is simply the time since birth; the period is the time or date of observation of an outcome; and the cohort is the time of birth. All of these are functions of time hence, time-varying. The idea here is that the variations in an outcome can be a function of all or any of age (changes in time since birth), period (changes due to specific events or time of observation) and cohort (differences in characteristics due to differences in times of birth) effects. As these can (simultaneously) impact an outcome. Besides having each of the variables as a linear function of the other creates a substantial identification issue – where we cannot attribute the effect of a variable to an outcome. In this study, I have used the phrase extensively 'an additional year of schooling explains…'. The problem with this can be material to the point where the interpretation of such becomes completely misleading. An 'additional year of schooling' for a respondent a year ago can be quite distinct from 'an additional year of schooling' thirty (or even two) years ago, as these differences in 'period' may impact the effects of the outcome of 'the year of schooling' in question. It is typical for social scientists to assess identification issues that accrue from these age-period-cohort effects. One simple step taken to assess this in simple regression analysis is to examine the effect of the younger cohort on the dependent variable of interest, to assess any variability in outcome. An example of this (not exactly) bias is seen in Appendix A1, see Tables A1.5 and A1.6. Outputs of Table A1.5 are very much like those of Table A1.6, the exception is that I have controlled for age in the former. The effect of this is that the reform indicator which is a function of year of birth (cohort) and age at first grade shows clear evidence that both the reform indicator and the age variables are confounded. A look at A1.6 relative A1.5 suggests a substantial difference in the mean effect of the reform indicator on schooling and, the effect of the reform indicator on skill (after controlling for age) becomes statistically insignificant. This suggests extreme care is needed to prevent or include certain variables (as certain covariates can result in such bias, just like the absence of some can create substantial omission bias in estimates). A closely related factor is the personality traits of respondents observed in adulthood as in this study (see the STEP data for Kenya). This presents endogeneity issues if the traits are absent, at the time of birth or if the variables/traits vary across the lifecycle. Observing these traits in adulthood may suggest the outcomes presented are not the sole effects of the traits but of the associated age, period, and cohort effects. However, I argue that the OLS estimates are sufficiently free from bias, particularly, bias from omitted variables. The work of Mosca and Wright (2018) concurs with the unbiasedness of the personality traits. Sadly, it is not enough to claim estimates are free from all biases on this basis of a study. The more recent work of Fitzenberger et al. (2022) suggests these traits are substantially heterogeneous across cohorts. Furthermore, Fitzenberger et al. (2022) showed that, across the lifecycle, personality traits are stable, but non-constant across age profiles with substantial differences across periods. Their study suggests useful estimates of the effects of personality traits require separating the age-period-cohort effects. Hence, I have highlighted this as a limitation of this study. However, inspired by the studies of Heckman et al., (2006; 2019) in a dynamic framework, using structural equations and the corresponding measurement equations where the latency of the traits (skills more broadly) is assumed (please, see Chapter 5) I show that findings from my OLS estimations are less bias if not completely free from bias, as the outcomes from the OLS estimates are consistent with the findings from the SEM (Structural Equation Modelling). This SEM approach is deemed to obtain consistent estimates of personality traits and cognitive skills (see Krishnakumar and Nogales, (2020)) by an assumption of their latency— the existence of the skills or traits from birth. It is important to note that it is only an ‘assumption’ in a structural estimation. Suggesting Fitzenberger et al. (2022) argument on the need to disentangle the age-cohort effect is a useful step for empirical works of this nature. Hence, I acknowledge the limitations of the estimates of the effects of these personality traits.

Using a repeated cross-section or a panel dataset with useful direct measures of skill to re-examine these estimates and further test the robustness of outcomes is strongly encouraged for future studies.

Instrument Validity is a major limitation across this study, particularly for External Returns (Chapter 4) where the OLS results are presented. In this study, a total of four instruments were used. Firstly, the reform dummy then the interaction of the reform dummy and each of the quarters of birth (Q2, Q3 and Q4). Please, see Table 2.1 for each of these. The use of these instruments was inspired by the study of Acemoglu (1999) where he estimated Private and Social Returns. I now turn to discuss the test outcomes. Appendix A3 contain all the results. To start with, In the initial analysis (chapter 2) where the outcome variable is standardised reading proficiency taking individual and average schooling as endogenous (Table A3.1, panel 1). The under-identification test uses the Anderson Canon correlation statistic, and it is used to assess if the instruments are less, compared to the endogenous variables. Here, the p-value is (weakly) significant at the 1% level, however, showing no problem of under-identification. Rejecting the null hypothesis ‘there is a problem of under-identification’. The next is the weak identification test, using the Cregg-Donald Wald F statistic, the Stock-Yogo weak identification test is used to assess if the IV fully define the endogenous variable. The findings suggest the test fails to hold even at the 25% critical value, the Wald F-statistic of 2.601 is lower than the 25% critical value of 7.910 suggesting the instruments are weak. The next is the Sargan Statistic for overidentification, this is to assess that the instruments are not correlated with the error terms. With a null hypothesis that the instruments are valid. Here, the Chi-Square value is not statistically significant, hence, the null hypothesis cannot be rejected. Hence, the instruments are not correlated with the error terms. Next, here the test of endogeneity of the treatment variables suggests a Chi-Square value that is weakly statistically significant, suggesting the treatment variables are endogenous and may not be treated as exogenous. Here, to test for heteroskedasticity, using the Pagan-Hall general test statistic, with a null hypothesis that the disturbance is homoscedastic, the evidence suggests the presence of heteroskedasticity. However, I have used clustered robust standard errors to account for this.

In treating average and individual schooling as endogenous with log-hourly earnings in USD as the outcome, please, see Panel 2 of Table A3.1. The under-identification test. Here, the p-value is statistically insignificant, showing the problem of under-identification. The next is the weak identification test, the Stock-Yogo weak identification test used to assess if the IV fully define the endogenous variable. The findings suggest the test holds weakly at the 25% critical value. The next is the Sargan Statistic for overidentification, this is to assess that the instruments are not correlated with the error terms. With a null hypothesis that the instruments are valid. Here, the Chi-Square value is statistically significant, hence, the null hypothesis is rejected. The instruments are correlated with the error terms. Next, the test for endogeneity of the treatment variables suggests, a Chi-Square value that is statistically significant at the 1% level suggesting the treatment variables are endogenous and may not be treated as exogenous. To test for heteroskedasticity, using the Pagan-Hall general test statistic, with a null hypothesis that the disturbance is homoscedastic, the evidence suggests the presence of heteroskedasticity. However, I have used clustered robust standard errors to account for this. This made it possible to draw inferences from the sample to the population.

Next, in treating average and individual skill as exogenous with log-hourly earnings in USD as the outcome, please, see Panel 3 of Table A3.1. Here, for the under-identification test, the p-value is statistically insignificant, showing the problem of under-identification. The next is the weak identification test, the Stock-Yogo weak identification test used to assess if the IV fully define the endogenous variable. The findings suggest the test does not hold at any critical values, suggesting weak identification. The next is the Sargan Statistic for overidentification, this is to assess that the instruments are not correlated with the error terms. With a null hypothesis that the instruments are valid. Here, the Chi-Square value is statistically significant at the 5% level, hence, the null hypothesis is rejected. The instruments are correlated with the error terms. Next, the test for endogeneity of the treatment variables suggests a Chi-Square value that is statistically significant at the 1% level suggesting the treatment variables are endogenous and may not be treated as exogenous. To test for heteroskedasticity, using the Pagan-Hall general test statistic, with a null hypothesis that the disturbance is homoscedastic, the evidence suggests the presence of heteroskedasticity. However, I have used clustered robust standard errors to account for this. This made it possible to draw inferences from the sample to the population.

Tables A3.2-A3.6 show substantially improved outcomes of the tests, particularly, after treating the individual schooling and skill as only endogenous variables for standardised skill and log earnings outcomes. The limitations in data extended to the issues that marred the outcomes of 2SLS-IV estimates, particularly, this suggests that the returns to external schooling and skill must be treated as exogenous. Interestingly, the OLS outcomes showed useful robustness. Please, Chapter 4. To further test the robustness of the outcomes of private returns to schooling and skills (Chapters 2 and 3), overcoming possible issues that mar the internal and external validity of estimates, inspired by the study of Krishnakumar and Nogales (2020), I deployed the technology of skills formation where the pecuniary and non-pecuniary private returns to schooling and skills were examined in a dynamic framework. Please, see Chapter 5 for details. However, as earlier emphasised an implication of these for future research is to further assess the robustness of the returns to external and social returns by obtaining useful instruments that will meet the tests of validity requirement and better still use a repeated cross-section or panel data to carry out similar analysis. This will further support policymaking by providing useful estimates from which causal inferences are drawn.

Appendices

Appendix A1: Tables of Additional Results

	YoS
p1985_	0.405
	(0.097)
father_educ_456	1.724***
	(0.000)
ses_1	-1.054***
	(0.000)
avg_skill_	0.011*
	(0.021)
apvlit_c	0.020***
	(0.000)
_cons	4.964***
	(0.000)
N	3145
adj. R-sq	0.34

Table A1.1: Effects of Aggregate Skill on Schooling. Please, refer to Table 2.11. Analysis and Discussions of Chapter 2.

	1	2	3	4	5	6	7	8	9	10
	Pool	Pool	father_educ_456	father_educ_456	father_educ_123	father_educ_123	ses_1	ses_1	ses_23	ses_23
	apvlit_d_3	apvlit_d_4	apvlit_d_3	apvlit_d_4	apvlit_d_3	apvlit_d_4	apvlit_d_3	apvlit_d_4	apvlit_d_3	apvlit_d_4
p1985_	0.193*	0.085	-0.144	0.116	0.234**	0.075	0.389*	0.632	0.222*	0.159
	(0.029)	(0.546)	(0.523)	(0.688)	(0.008)	(0.626)	(0.016)	(0.107)	(0.027)	(0.297)
father_educ_456	0.467***	0.721***								
	(0.000)	(0.000)								
ses_1	-0.244***	-0.136								
	(0.000)	(0.186)								
stratum_N	-0.008	0.263	0.009	0.0912	0.011	0.375*	-0.214	-0.019	0.091	0.36
	(0.944)	(0.107)	(0.961)	(0.638)	(0.928)	(0.025)	(0.133)	(0.939)	(0.500)	(0.063)
stratum_L	0.008	0.363*	0.189	0.27	-0.033	0.439**	-0.172	0.075	0.098	0.470*
	(0.936)	(0.026)	(0.43)	(0.215)	(0.724)	(0.008)	(0.242)	(0.783)	(0.399)	(0.028)
stratum_M	0.089	0.294	0.289	0.277	0.0454	0.304	-0.124	0.39	0.217*	0.361
	(0.255)	(0.068)	(0.119)	(0.22)	(0.578)	(0.054)	(0.324)	(0.072)	(0.032)	(0.058)
_cons	-0.751***	-1.986***	-0.24	-1.379***	-0.797***	-1.962***	-0.806***	-2.374***	-0.803***	-1.904***
	(0.000)	(0.000)	(0.459))	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	3123	3123	650	650	2484	2484	731	731	2403	2403
adj. R-sq										

Table A1.2: Please, refer to Table 2.11. Analysis and Discussions of Chapter 2. Probit models, regressing dummies of categories of schooling on the reform indicator, accounting for background characteristics, district-, and strata-specific effects, following similar specifications of Equations 2.5 and 2.6 as in (6) and (14) of Table 2.11.

	1	2	3	4	5	6	7	8	9	10
	Pool	Pool	father_educ_456		father_educ_123		ses_1		ses_23	
	isced34A	isced56	isced34A	isced56	isced34A	isced56	isced34A	isced56	isced34A	isced56
p1985_	0.377***	-0.231*	1.051**	-0.640*	0.349***	-0.177	0.268	0.419	0.423***	-0.151
	(0.000)	(0.037)	(0.01)	(0.027)	(0.000)	(0.117)	(0.099)	(0.087)	(0.000)	(0.224)
father_educ_456	-0.013	0.874***								
	(0.773)	(0.000)								
ses_1	-0.262***	-0.17								
	(0.000)	(0.058)								
stratum_N	0.567***	0.060	0.533***	-0.071	0.589***	0.126	0.438**	-0.059	0.604***	0.167
	(0.000)	(0.762)	(0.000)	(0.789)	(0.000)	(0.482)	(0.009)	(0.812)	(0.000)	(0.515)
stratum_L	0.625***	-0.047	0.638***	-0.132	0.619***	-0.004	0.705***	-0.352	0.609***	0.109
	(0.000)	(0.798)	(0.000)	(0.609)	(0.000)	(0.984)	(0.000)	(0.161)	(0.000)	(0.621)
stratum_M	0.693***	0.0758	0.744***	-0.028	0.668***	0.124	0.776***	-0.169	0.663***	0.264
	(0.000)	(0.658)	(0.000)	(0.912)	(0.000)	(0.456)	(0.000)	(0.489)	(0.000)	(0.174)
_cons	-1.290***	-1.255***	-1.942***	-0.013	-1.329***	-1.320***	-1.479***	-1.565***	-1.320***	-1.222***
	(0.000)	(0.000)	(0.000)	(0.973)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	3147	3147	658	658	2500	2500	734	734	2424	2424
adj. R-sq										

Notes: Table A1.3: Please, refer to Table 2.11. Analysis and Discussions of Chapter 2. Probit models, regressing dummies of categories of skills on the reform indicator, accounting for background characteristics, district-, and strata-specific effects, following similar specifications of Equations 2.5 and 2.6 as in (6) and (14) of Table 2.11.

	1	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Having a father with Post-Sec Ed				Having a mother with Post-Sec Ed				SES at Age 15			
Indicator	1		0		1		0		1 (Low)		3 (High).	
	Schooling	Skill	Schooling	Skill	Schooling	Skill	Schooling	Skill	Schooling	Skill	Schooling	Skill
Reform Indicator	0.696**	0.198***	-1.792*	-0.036	0.870***	0.244***	-2.013*	0.064	0.628*	0.255***	1.148***	0.307***
	(0.001)	(0.000)	(0.024)	(0.862)	(0.000)	(0.000)	(0.039)	(0.801)	(0.013)	(0.000)	(0.000)	(0.000)
Father Post-Sec Ed	5.451***	0.883***										
	(0.000)	(0.000)										
Reform × Father Post-Sec	-2.488**	-0.233										
	(0.003)	(0.275)										
Father Under-Sec Ed			-5.451***	-0.883***								
			(0.000)	(0.000)								
Reform × Father Under-Sec			2.488**	0.233								
			(0.003)	(0.275)								
Mother Post-Sec Ed					5.866***	0.739**						
					(0.000)	(0.004)						
Reform × Mother Post-Sec Ed					-2.882**	-0.180						
					(0.004)	(0.487)						
Mother Under-Sec Ed							-5.866***	-0.739**				
							(0.000)	(0.004)				
Reform x Mother Unde-Sec Ed							2.882**	0.180				
							(0.004)	(0.487)				
Low SES at Age 15									-2.732***	-0.383***		
									(0.000)	(0.001)		
Reform × Low SES									0.845	0.0170		
									(0.076)	(0.890)		
High SES at Age 15											2.819***	0.485**
											(0.000)	(0.001)
Reform × High SES											-1.119	-0.163
											(0.078)	(0.316)
Nairobi_Strata	0.636***	-0.303***	0.636***	-0.303***	0.718***	-0.279***	0.718***	-0.279***	0.827***	-0.259***	0.960***	-0.234***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Large_Strata	0.581**	-0.255***	0.581**	-0.255***	0.678***	-0.232***	0.678***	-0.232***	0.787***	-0.214***	0.795***	-0.212***
	(0.002)	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Medium_Strata	0.817***	-0.224***	0.817***	-0.224***	0.951***	-0.190***	0.951***	-0.190***	1.128***	-0.159**	1.229***	-0.139**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.005)
Small_Strata (Reference)												
_cons	9.214***	0.008	14.67***	0.891***	9.236***	0.0143	15.10***	0.753**	10.18***	0.140*	8.980***	-0.044
	(0.000)	(0.889)	(0.000)	(0.000)	(0.000)	(0.805)	(0.000)	(0.003)	(0.000)	(0.033)	(0.000)	(0.475)
N	3156	3156	3156	3156	3156	3156	3156	3156	3145	3145	3145	3145
R-sq	0.130	0.092	0.130	0.092	0.095	0.052	0.095	0.052	0.071	0.042	0.045	0.028
adj. R-sq	0.128	0.090	0.128	0.090	0.093	0.050	0.093	0.050	0.070	0.040	0.043	0.026

Note: Table A1.4 reports outputs of variants of models 2.5; and 2.6 as columns (1), (3), (5), (7), (9), and (11); and (2), (4), (6), (8), (10) and (12) respectively, following the Difference-in-Differences techniques. The outcome of the former is schooling as actual years of schooling and the latter skills, as standardised reading proficiency. Both models assess the effect of the indicator of the 1985 curriculum reform, p1985, and background characteristics (as indicators of the father's post-secondary education; mother's post-secondary education; and socioeconomic status, (with mid-level socioeconomic status as the reference category here)). Controls include strata-specific effects which is a basis of the stratified sampling (based on the number of households across cities) across Nairobi, strata_N, other large cities, strata_L with over 100 000HH, medium cities, strata_M with over 60 000HH but under 100 000HH, and other cities under, strata_S with under 60 000HH (reference category). The p-values in parentheses: * $p<0.05$; ** $p<0.01$; *** $p<0.001$. **With no clustering at the district level**. Here, also shown are the direct reverse effects of the treated and untreated – e.g., the effects of having father with post-secondary education compared to having a father with 'under' post-secondary education.

	1	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Outcome Variables	Years of Education (Measure of Educational Attainment) _actual number of years of schooling						Reading Proficiency (Measure of Cognitive Skill) _standardised average plausible values.						
Reform Indicator	1.107***	1.142***	1.152***	1.180***	1.131***	1.151***	-0.001	0.006	0.011	0.016	0.006	0.009	-0.137*
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.995)	(0.936)	(0.882)	(0.824)	(0.929)	(0.901)	(0.033)
Age	0.021	0.014	0.010	0.006	0.023*	0.024*	-0.009*	-0.011**	-0.012***	-0.012***	-0.009*	-0.009*	-0.012***
	(0.053)	(0.193)	(0.369)	(0.588)	(0.028)	(0.026)	(0.011)	(0.001)	(0.000)	(0.000)	(0.012)	(0.012)	(0.000)
Gender	-0.831***	-0.831***	-0.869***	-0.863***	-0.840***	-0.832***	-0.177***	-0.178***	-0.185***	-0.184***	-0.179***	-0.178***	-0.074*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.018)
Father Post-Sec Ed	2.777***				2.531***	1.990***	0.593***				0.547***	0.512***	0.259***
	(0.000)				(0.000)	(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
Mother Post-Sec Ed		2.707***				1.020***		0.463***				0.052	-0.074
		(0.000)				(0.000)		(0.000)				(0.277)	(0.132)
Low SES at Age 15			-1.825***		-1.461***	-1.407***			-0.326***		-0.247***	-0.239***	-0.057
			(0.000)		(0.000)	(0.000)			(0.000)		(0.000)	(0.000)	(0.229)
High SES at Age 15				1.499***		0.384				0.277***		0.065	0.018
				(0.000)		(0.086)				(0.000)		(0.206)	(0.693)
Priv_Sch	1.238***	1.227***	1.355***	1.362***	1.200***	1.149***	0.086	0.092	0.114	0.116	0.081	0.078	-0.070
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.217)	(0.163)	(0.112)	(0.100)	(0.247)	(0.265)	(0.331)
Other_Sch	0.402	0.383	0.348	0.409	0.287	0.246	0.061	0.063	0.062	0.071	0.047	0.043	0.013
	(0.060)	(0.067)	(0.134)	(0.063)	(0.176)	(0.225)	(0.309)	(0.296)	(0.331)	(0.243)	(0.435)	(0.471)	(0.794)
AnotherCity_SchLoc	-0.929***	-1.025***	-1.197***	-1.229***	-0.870***	-0.817***	-0.246***	-0.279***	-0.308***	-0.312***	-0.236***	-0.232***	-0.134***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ForeignCity_SchLoc	0.514	0.978	1.250	0.951	0.621	0.604	-0.496	-0.388	-0.360	-0.414	-0.495*	-0.501*	-0.566***
	(0.619)	(0.356)	(0.207)	(0.379)	(0.491)	(0.488)	(0.059)	(0.141)	(0.140)	(0.111)	(0.038)	(0.034)	(0.001)
Nairobi_Strata	0.979*	1.069**	1.175**	1.286**	0.946*	0.945*	-0.239**	-0.211*	-0.190*	-0.171	-0.241**	-0.239**	-0.403***
	(0.010)	(0.009)	(0.006)	(0.003)	(0.011)	(0.010)	(0.006)	(0.023)	(0.039)	(0.062)	(0.004)	(0.004)	(0.000)
Large_Strata	0.854*	0.951*	1.065*	1.071*	0.865*	0.862*	-0.207*	-0.180*	-0.162	-0.162	-0.206*	-0.206*	-0.359***
	(0.020)	(0.016)	(0.011)	(0.010)	(0.018)	(0.017)	(0.014)	(0.048)	(0.084)	(0.083)	(0.012)	(0.012)	(0.000)
Medium_Strata	1.128**	1.258***	1.410***	1.507***	1.112**	1.112**	-0.182*	-0.145	-0.118	-0.100	-0.183*	-0.180*	-0.364***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.019)	(0.079)	(0.182)	(0.251)	(0.017)	(0.017)	(0.000)
Avg. Yos (district)													0.108
													(0.315)
Avg. Yos Squ (district)													-0.00476
													(0.335)
years_educ_act													0.125***
													(0.000)
_cons	8.894***	9.291***	10.16***	9.609***	9.177***	9.054***	0.653***	0.766***	0.911***	0.809***	0.697***	0.684***	-1.002
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.084)
N	3121	3121	3121	3121	3121	3121	3121	3121	3121	3121	3121	3121	3121
R-sq	0.168	0.136	0.126	0.100	0.193	0.200	0.126	0.090	0.088	0.075	0.137	0.138	0.332
adj. R-sq	0.165	0.133	0.122	0.096	0.190	0.196	0.122	0.087	0.084	0.072	0.134	0.134	0.328

Note: Table A1.5 reports outputs of variants of models 2.5; and 2.6 as columns (1) - (6); and columns (7) - (13) respectively. The outcome of the former is schooling as actual years of schooling, and the outcome of the latter, is skill, as standardised reading proficiency. Both models assess the effects of the indicator of the 1985 curriculum reform, p1985, and background characteristics (as indicators of the father's post-secondary education; mother's post-secondary education; and socioeconomic status, (with mid-level socioeconomic status as the reference category here)). Also accounted for are the effects of type and location of school or institution attended for the most recent/final qualification. The school types accounted for are private and other (this includes homeschooling) with the reference category as public-funded schools. For school location, with 'other cities' and those that studied in foreign cities (e.g., in Uganda) accounted for with those that studied in the same city as their current residence as the reference category. The use of the latter is to possibly capture the effects of migration, particularly, rural-urban drifts within Kenya, acknowledging the urban analytical sample of this study. Other controls include strata-specific effects which is a basis of the stratified sampling (based on the number of households across cities) across Nairobi, strata_N, other large cities, strata_L with over 100 000HH, medium cities, strata_M with over 60 000HH but under 100 000HH, and other cities under, strata_S with under 60 000HH (reference category). The p-values in parentheses: * p<0.05; ** p<0.01; *** p<0.001. With clustered robust standard error (at the district level). Also accounted for, are the effects of Age and Gender.

	1	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Outcome Variables	Years of Education (Measure of Educational Attainment) _actual number of years of schooling						Reading Proficiency (Measure of Cognitive Skill) _standardised average plausible values.						
Reform Indicator	0.632*	0.827**	0.933**	1.047***	0.596*	0.607*	0.206***	0.258***	0.277***	0.298***	0.202***	0.204***	0.130*
	(0.023)	(0.004)	(0.002)	(0.000)	(0.029)	(0.024)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.019)
Gender	-0.840***	-0.837***	-0.873***	-0.865***	-0.850***	-0.842***	-0.173***	-0.174***	-0.180***	-0.179***	-0.175***	-0.175***	-0.070*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.019)
Father Post-Sec Ed	2.746***				2.498***	1.960***	0.606***				0.559***	0.523***	0.275***
	(0.000)				(0.000)	(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
Mother Post-Sec Ed		2.689***				1.011***		0.477***				0.055	-0.069
		(0.000)				(0.000)		(0.000)				(0.247)	(0.164)
Low SES at Age 15			-1.820***		-1.452***	-1.397***			-0.333***		-0.250***	-0.242***	-0.063
			(0.000)		(0.000)	(0.000)			(0.000)		(0.000)	(0.000)	(0.197)
High SES at Age 15				1.498***		0.390				0.280***		0.063	0.015
				(0.000)		(0.082)				(0.000)		(0.214)	(0.732)
Priv_Sch	1.230***	1.221***	1.351***	1.360***	1.191***	1.140***	0.090	0.096	0.120	0.121	0.085	0.081	-0.062
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.194)	(0.138)	(0.095)	(0.081)	(0.225)	(0.242)	(0.363)
Other_Sch	0.406	0.385	0.350	0.410	0.292	0.251	0.060	0.062	0.0603	0.070	0.046	0.042	0.012
	(0.061)	(0.068)	(0.135)	(0.064)	(0.175)	(0.223)	(0.321)	(0.309)	(0.344)	(0.252)	(0.452)	(0.489)	(0.826)
AnotherCitySchLoc	-0.915***	-1.014***	-1.189***	-1.224***	-0.855***	-0.801***	-0.252***	-0.288***	-0.318***	-0.323***	-0.242***	-0.238***	-0.143***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ForeignCitySchLoc	0.544	0.995	1.261	0.958	0.655	0.637	-0.509*	-0.401	-0.373	-0.429	-0.508*	-0.513*	-0.584***
	(0.603)	(0.352)	(0.206)	(0.377)	(0.475)	(0.472)	(0.048)	(0.118)	(0.115)	(0.089)	(0.029)	(0.026)	(0.000)
Nairobi_Strata	0.959*	1.055**	1.165**	1.280**	0.925*	0.924*	-0.230**	-0.199*	-0.178	-0.158	-0.233**	-0.231**	-0.385***
	(0.011)	(0.009)	(0.006)	(0.002)	(0.012)	(0.011)	(0.007)	(0.030)	(0.052)	(0.082)	(0.005)	(0.005)	(0.000)
Large_Strata	0.829*	0.933*	1.052*	1.063*	0.837*	0.834*	-0.196*	-0.166	-0.147	-0.146	-0.196*	-0.196*	-0.339***
	(0.024)	(0.017)	(0.011)	(0.010)	(0.022)	(0.021)	(0.016)	(0.059)	(0.108)	(0.107)	(0.014)	(0.013)	(0.000)
Medium_Strata	1.103**	1.240***	1.397***	1.499***	1.085**	1.084**	-0.171*	-0.131	-0.102	-0.0833	-0.172*	-0.170*	-0.344***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.002)	(0.001)	(0.024)	(0.108)	(0.243)	(0.335)	(0.022)	(0.022)	(0.000)
Avg. Yos (district)													0.091
													(0.389)
Avg. Yos Squ (district)													-0.004
													(0.415)
years_educ_act													0.124***
													(0.000)
_cons	9.946***	9.984***	10.64***	9.901***	10.36***	10.26***	0.197**	0.212**	0.329***	0.192**	0.266***	0.253***	-1.497**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.001)	(0.000)	(0.006)	(0.000)	(0.000)	(0.007)
N	3121	3121	3121	3121	3121	3121	3121	3121	3121	3121	3121	3121	3121
R-sq	0.167	0.136	0.125	0.100	0.192	0.198	0.122	0.085	0.082	0.069	0.134	0.135	0.326
adj. R-sq	0.164	0.133	0.122	0.097	0.189	0.195	0.119	0.082	0.079	0.066	0.131	0.131	0.323

Note: Table A1.6 reports outputs of variants of models 2.5; and 2.6 as columns (1) - (6); and columns (7) - (13) respectively. The outcome of the former is schooling as actual years of schooling, and the outcome of the latter is skills, as standardised reading proficiency. Both models assess the effects of the indicator of the 1985 curriculum structural reform, p1985_and background characteristics (as indicators of the father's post-secondary education; mother's post-secondary education; and socioeconomic status, (with mid-level socioeconomic status as the reference category here)). Also accounted for are the effects of type and location of school or institution attended for the most recent/final qualification. The school type accounted for are private and other (this includes homeschooling) with the reference category as public-funded schools. For school location, accounted for are other cities within Kenya, and those that studied in foreign cities (e.g., in Uganda) with those that study in the same city as their current residence as the reference category. The use of the latter is to possibly capture the effects of migration, particularly, rural-urban drifts within Kenya, acknowledging the urban analytical sample of this study. Other controls include strata-specific effects which is a basis of the stratified sampling (based on the number of households across cities) across Nairobi, strata_N, other large cities, strata_L with over 100 000HH, medium cities, strata_M with over 60 000HH but under 100 000HH, and other cities under, strata_S with under 60 000HH (reference category). The p-values in parentheses: * $p<0.05$; ** $p<0.01$; *** $p<0.001$. With clustered robust standard error (at the district level). Here, in lieu of Age and Gender, only gender is accounted for.

	Indicator of Mother's Post–Sec Educ.			Indicator of Father's Post-Sec Educ.			Indicator of Socioeconomic Status			
school_type	0	1	total (freq)	0	1	total (freq)	1	2	3	total (freq)
Public	2285	272	2557	2059	498	2557	632	1656	258	2546
private	179	46	225	158	67	225	44	146	35	225
Others	324	65	389	290	99	389	60	257	71	388
total	**2788**	**383**	**3171**	**2507**	**664**	**3171**	**736**	**2059**	**364**	**3159**
school_location										
Samecity	1011	232	1243	883	360	1243	232	814	194	1240
anothercity_province	1736	137	1873	1595	278	1873	492	1218	155	1865
another_country	34	12	46	22	24	46	11	21	13	45
total	**2781**	**381**	**3162**	**2500**	**662**	**3162**	**735**	**2053**	**362**	**3150**
ISCED										
no qualification	344	6	350	337	13	350	152	171	27	350
completion of primary education	675	19	694	655	39	694	233	415	44	692
completion of lower sec education	411	25	436	389	47	436	100	288	47	435
completion of sec and some post-sec education	1014	143	1157	892	265	1157	214	808	127	1149
completion of (adv) post-sec non-tertiary educ.	241	68	309	184	125	309	36	227	46	309
completion of tertiary education	212	127	339	163	176	339	47	207	84	338
Total	**2897**	**388**	**3285**	**2620**	**665**	**3285**	**782**	**2116**	**375**	**3273**

Note: Table A1.7.1 reports the descriptive evidence (frequencies) of the type of schooling (private, public and others (including homeschooling)) received; and the location of the most recent school/institution attended, and the category or level of schooling obtained (*isced*). Variables are disaggregated by background characteristics: Indicators of Father/Mother Post-Secondary Schooling (1 indicates having a father/mother with post-secondary schooling; and 0 indicates otherwise). For the indicator of socioeconomic status (1 indicates low SES; 2 indicates mid-level SES; and 3 indicates high SES).

	Indicator of Mother’s Post–Sec Educ.		Indicator of Father’s Post-Sec Educ.		Indicator of Socioeconomic Status		
school_type	0	1	0	1	1	2	3
Public	81.96%	71.02%	82.13%	75%	85.87%	80.43%	70.88%
private	6.42%	12.01%	6.30%	10.09%	5.98%	7.09%	9.62%
Others	11.62%	16.97%	11.57%	14.91%	8.15%	12.48%	19.51%
total	100	100	100	100	100	100	100
school_location							
Samecity	36.35%	60.89%	35.32%	54.38%	31.56%	39.65%	53.59%
anothercity_province	62.42%	35.96%	63.8%	41.99%	66.94%	59.33%	42.82%
another_country	1.22%	3.15%	0.88%	3.63%	1.50%	1.02%	3.59%
Total	100	100	100	100	100	100	100
ISCED							
no qualification	11.87%	1.55%	12.86%	1.95%	19.44%	8.08%	7.20%
completion of primary education	23.30%	4.90%	25.00%	5.86%	29.80%	19.61%	11.73%
completion of lower sec education	14.19%	6.44%	14.85%	7.07%	12.79%	13.61%	12.53%
completion of sec and some post-sec education	35.00%	36.86%	34.05%	39.85%	27.37%	38.19%	33.87%
completion of (adv) post-sec non-tertiary education.	8.32%	17.53%	7.02%	18.80%	4.60%	10.73%	12.27%
completion of tertiary education	7.32%	32.73%	6.22%	26.47%	6.01%	9.78%	22.40%
Total	**100**	**100**	**100**	**100**	**100**	**100**	**100**

Note: Table A1.7.2 reports the descriptive evidence (percentages) of the type of schooling (private, public, and others (including homeschooling)) received; the location of the most recent school/institution attended, and the category or level of schooling obtained (*isced*). Variables are disaggregated by, background characteristics: Indicators of Father/Mother Post-Secondary Schooling (1 indicates having a father/mother with post-secondary schooling; 0 indicates otherwise). For the indicator of socioeconomic status (1 indicates low SES; 2 indicates mid-level SES; and 3 indicates high SES). All figures are in percentages of the total percentage across columns (adds up to 100% across columns).

Age_Start	Freq	Percentage	Cumulative %
3	5	0.16	0.16
4	22	0.69	0.85
5	252	7.93	8.78
6	1,105	34.79	43.58
7	1,362	42.88	86.46
8	316	9.95	96.41
9	63	1.98	98.39
10	34	1.07	99.46
11	5	0.16	99.62
12	8	0.25	99.87
13	3	0.09	99.97
18	1	0.03	100
Total	**3,176**	**100**	

Note: Table A1.8 reports the descriptive evidence (frequencies and percentages) of age at first grade in Kenya based on the main analytical sample.

Age_Start	Freq	Percentage	Cumulative %
5	252	8.01	8.01
6	1,105	35.10	43.11
7	1,362	43.27	86.37
8	316	10.04	96.41
9	63	2.00	98.41
10	34	1.08	99.49
11	5	0.16	99.65
12	8	0.25	99.90
13	3	0.10	100
Total	**3,148**	**100**	

Note: Table A1.9 reports the descriptive evidence (frequencies and percentages) of age at first grade in Kenya based on the TRIMMED analytical sample. Dropping age_start under 5 and over 13.

	1	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Outcome Variables	Years of Education (Measure of Educational Attainment) _actual number of years of schooling						Reading Proficiency (Measure of Cognitive Skill) _standardised average plausible values.						
Reform Indicator	0.640*	0.834**	0.939**	1.054***	0.605*	0.617*	0.205**	0.256***	0.276***	0.297***	0.201**	0.203**	0.128*
	(0.023)	(0.004)	(0.002)	(0.000)	(0.029)	(0.024)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.023)
Gender	-0.858***	-0.850***	-0.892***	-0.882***	-0.869***	-0.860***	-0.176***	-0.176***	-0.183***	-0.181***	-0.178***	-0.178***	-0.0711*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.018)
Father Post-Sec Ed	2.743***				2.489***	1.949***	0.606***				0.556***	0.518***	0.275***
	(0.000)				(0.000)	(0.000)	(0.000)				(0.000)	(0.000)	(0.000)
Mother Post-Sec Ed		2.684***				1.011***		0.479***				0.0586	-0.068
		(0.000)				(0.000)		(0.000)				(0.226)	(0.167)
Low SES at Age 15			-1.828***		-1.452***	-1.397***			-0.342***		-0.258***	-0.249***	-0.0676
			(0.000)		(0.000)	(0.000)			(0.000)		(0.000)	(0.000)	(0.171)
High SES at Age 15				1.506***		0.404				0.288***		0.0698	0.0201
				(0.000)		(0.073)				(0.000)		(0.168)	(0.650)
Priv_Sch	1.254***	1.227***	1.361***	1.378***	1.211***	1.154***	0.11	0.112	0.137	0.140*	0.104	0.099	-0.0456
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.104)	(0.081)	(0.058)	(0.045)	(0.133)	(0.147)	(0.498)
Other_Sch	0.407	0.389	0.363	0.417	0.297	0.255	0.0604	0.0637	0.0633	0.072	0.0469	0.0423	0.0116
	(0.062)	(0.066)	(0.121)	(0.059)	(0.170)	(0.218)	(0.317)	(0.295)	(0.321)	(0.240)	(0.442)	(0.483)	(0.824)
AnotherCitySchLoc	-0.910***	-1.003***	-1.174***	-1.213***	-0.846***	-0.791***	-0.250***	-0.285***	-0.314***	-0.320***	-0.240***	-0.235***	-0.140***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ForeignCitySchLoc	0.537	0.992	1.26	0.952	0.653	0.635	-0.511*	-0.403	-0.373	-0.431	-0.508*	-0.514*	-0.585***
	(0.608)	(0.353)	(0.206)	(0.380)	(0.477)	(0.474)	(0.047)	(0.117)	(0.114)	(0.087)	(0.029)	(0.025)	(0.000)
Nairobi_Strata	0.966*	1.065**	1.181**	1.291**	0.936*	0.937*	-0.232**	-0.201*	-0.179*	-0.159	-0.234**	-0.232**	-0.387***
	(0.011)	(0.008)	(0.005)	(0.002)	(0.011)	(0.010)	(0.006)	(0.025)	(0.045)	(0.074)	(0.004)	(0.004)	(0.000)
Large_Strata	0.814*	0.921*	1.045*	1.054*	0.824*	0.821*	-0.196*	-0.165	-0.145	-0.144	-0.195*	-0.195*	-0.337***
	(0.025)	(0.018)	(0.011)	(0.010)	(0.023)	(0.022)	(0.015)	(0.057)	(0.109)	(0.108)	(0.013)	(0.013)	(0.000)
Medium_Strata	1.100**	1.243***	1.400***	1.499***	1.086**	1.087**	-0.175*	-0.134	-0.105	-0.0866	-0.176*	-0.173*	-0.349***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.019)	(0.092)	(0.219)	(0.304)	(0.017)	(0.017)	(0.000)
Avg. Yos (district)													0.0919
													(0.382)
Avg. Yos Squ (district)													-0.00397
													(0.411)
years_educ_act													0.124***
													(0.000)
_cons	9.957***	9.988***	10.64***	9.903***	10.37***	10.26***	0.200**	0.213***	0.333***	0.192**	0.271***	0.256***	-1.500**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.004)	(0.000)	(0.000)	(0.007)
N	3094	3094	3094	3094	3094	3094	3094	3094	3094	3094	3094	3094	3094
R-sq	0.167	0.136	0.126	0.101	0.192	0.199	0.122	0.085	0.084	0.07	0.135	0.136	0.327
adj. R-sq	0.165	0.133	0.124	0.098	0.19	0.195	0.12	0.082	0.081	0.067	0.132	0.132	0.323

Note: Table A1.10 reports the same output as Table A1.6 assessing the sensitivity of the age_start variable (after trimming as in Table A1.9) on the reform indicator (as a function of the age_start variable) and other variables on Table A1.6.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	ln_earnings_h_usd							
apvlit_c	0.004***		0.004***		0.004***		0.004***	
	(0.000)		(0.000)		(0.000)		(0.000)	
years_educ_act		0.121***		0.122***		0.116***		0.117***
		(0.000)		(0.000)		(0.000)		(0.000)
avg_skill	0.004		0.006***		0.006		0.005***	
	(0.626)		(0.000)		(0.473)		(0.000)	
avg_skill_sq	0.000004				-0.000004			
	(0.869)				(0.882)			
avg_yos		-0.249*		0.028		-0.258*		0.016
		(0.028)		(0.094)		(0.022)		(0.360)
avg_yos_sq		0.014*				0.014*		
		(0.013)				(0.014)		
tenure	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.00146***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
age	0.006	0.010***	0.006	0.011***	0.007*	0.010***	0.007*	0.0107***
	(0.050)	(0.001)	(0.050)	(0.000)	(0.018)	(0.001)	(0.017)	(0.000)
gender	-0.148**	-0.135**	-0.148**	-0.138**	-0.120*	-0.116*	-0.120*	-0.119*
	(0.003)	(0.004)	(0.003)	(0.004)	(0.015)	(0.014)	(0.015)	(0.012)
bmi	0.018**	0.016**	0.018**	0.016**	0.018**	0.016**	0.018**	0.0156**
	(0.003)	(0.005)	(0.003)	(0.006)	(0.002)	(0.005)	(0.002)	(0.006)
extraversion_av					0.089*	0.056	0.089*	0.048
					(0.037)	(0.180)	(0.038)	(0.249)
conscientious_avg					0.080	0.100*	0.080	0.100*
					(0.114)	(0.039)	(0.115)	(0.039)
openness_av					0.244***	0.180***	0.244***	0.183***
					(0.000)	(0.000)	(0.000)	(0.000)
stability_av					0.041	0.00004	0.0412	0.0003
					(0.401)	(0.999)	(0.402)	(0.996)
agreeableness_av					-0.049	-0.027	-0.049	-0.020
					(0.280)	(0.532)	(0.277)	(0.641)
_cons	-1.631*	-0.367	-1.757***	-1.762***	-2.972***	-1.168	-2.858***	-2.550***
	(0.042)	(0.548)	(0.000)	(0.000)	(0.000)	(0.070)	(0.000)	(0.000)
N	1756	1756	1756	1756	1754	1754	1754	1754
R-sq	0.143	0.220	0.143	0.217	0.165	0.233	0.165	0.231
adj. R-sq	0.140	0.217	0.140	0.215	0.159	0.228	0.160	0.226

Note: Table A1.11 reports the same output as Tables 3.8 and 3.17 assessing the choice of the variable specifications of average years of schooling (avg_yos) and average skills in districts of Kenya. This Table shows that average schooling enters as a quadratic, however average skill enters as a linear variable. Columns (1)-(4) are based on the specification of columns (9) of Tables 3.8 and 3.17. Columns (5)-(8) are based on the specifications of column (13) of Tables 3.8 and 3.17.

Appendix A2 – Foundation of Mincerian Earnings Function & the Big 5

A2.1: The Mincerian Earnings Function and Theoretical Foundations –

To estimate the returns (earnings) to education in Kenya, the empirical framework followed in this analysis is the Mincerian basic earnings function (Mincer, 1974). Amidst all criticisms, the Mincerian basic earning function has maintained its prominence as a baseline model for estimating returns to education. Developed by Jacob Mincer in 1974, the Mincerian earning function is underpinned by the human capital theory pioneered by the works of Gary Becker (1964) and Theodore Schultz (1962). The study of the latter, *"Reflections on the Investment in Man"* and the birth of a new literature that inspired the works of Gary Becker (1964) on human capital theory, two years later. Following the account of Harmon and Walker (2001), I will now discuss the theoretical foundations of the Mincerian earnings (wage) equation to set the scene for this analysis.

Central to the Mincerian basic earnings function, is the human capital theory (Becker 1964), with specifications:

$$\sum_{t=1}^{\mathcal{T}-s} \frac{\mathcal{W}_s - \mathcal{W}_{s-1}}{(1+r_s)^t} = \mathcal{W}_{s-1} + c_s \quad \text{..........} \quad (1)$$

This is the equilibrium point. Equilibrium is reached when the net present value or return to schooling is 0 –

Where *s* is schooling (or education measure), assumed to be infinitely divisible and the choice of *s* for a rational individual will be influenced by their expected PV (present value) of all future inflow (or income) to their retirement date, $\mathcal{T}$, net of c_s - their cost of schooling (or education). Where, $\mathcal{W}_s$ is the wage with schooling and $\mathcal{W}_{s-1}$, is the wage without schooling. Hence, the *s* where the present value of schooling, $\sum_{t=1}^{\mathcal{T}-s} \frac{\mathcal{W}_s - \mathcal{W}_{s-1}}{(1+r_s)^t}$ is equal to the cost of schooling, $\mathcal{W}_{s-1} + c_s$, is the optimum level of *s*. It is important to note that the expression, $\mathcal{W}_{s-1} + c_s$ represents the economic cost of schooling in PV–present value terms. It is the sum of the implicit or opportunity costs, $\mathcal{W}_{s-1}$ which is the current wage foregone; and the explicit or current costs of schooling, c_s, which is also known as the accounting or financial costs of schooling. The expression, $\sum_{t=1}^{\mathcal{T}-s} \frac{\mathcal{W}_s - \mathcal{W}_{s-1}}{(1+r_s)^t}$ representing the present value of schooling is the discounted sum of the future (after schooling to retirement) wage difference. The wage difference is due to schooling. Hence, at equilibrium, value = costs, or the net value, value – cost = 0. Therefore (1) can be expressed as the following:

$$\sum_{t=1}^{\mathcal{T}-s} \frac{\mathcal{W}_s - \mathcal{W}_{s-1}}{(1+r_s)^t} - (\mathcal{W}_{s-1} + c_s) = 0 \quad \text{.......} \quad (2)$$

Hence, at optimal schooling, the net present value is 0. Where r_s, is the discount rate, also known as the internal rate of return. As earlier stated, s – schooling (or education), is infinitely divisible. Knowing that the internal rate of return is the rate at which the net present value of schooling is 0, the optimal level of schooling (level of schooling where net present value is 0) and indeed the choice of educational attainment or educational investment decision is based on alternative uses of scarce resource for the rational man. Hence, investment in a level of schooling may require an internal rate of return, $r_s > i$, the market rate of return.

From (1), if $\mathcal{T}$, the date of the retirement is large, then the expression, $\sum_{t=1}^{\mathcal{T}-s} \frac{\mathcal{W}_s - \mathcal{W}_{s-1}}{(1+r_s)^t}$ is approximated such that, $\frac{\mathcal{W}_s - \mathcal{W}_{s-1}}{r_s} = \mathcal{W}_{s-1} + c_s$.......(3) with a sufficiently small c_s, explicit cost of schooling, (3) can be rearranged as thus:

$$r_s \approx \frac{\mathcal{W}_s - \mathcal{W}_{s-1}}{\mathcal{W}_s} \approx \log \mathcal{W}_s - \log \mathcal{W}_{s-1} \quad \text{.......} \quad (4)$$

Hence, from (4), the return, r_s, to the s^{th} level of schooling is approximately the difference (in log wages) of leaving at level s and a period of lag of level s, which is the level, s-1. It is important to note that the *level*, as used here, can be any fixed intervals representing (typically), the years of schooling (Harmon and Walker, 2001). The Mincerian earnings function has a well-defined underpinning to the human capital theory specified by Becker (1964) as above. Specifically, in the empirical model specified by Mincer (1974), the return variable, ρ_s is not only deemed to be the proportionate effect on wages owing to an increase in s but it is also deemed to be the private financial return to education (or schooling) acknowledging assumptions of the human capital theory by Becker (1964), specifically, the 'sufficiently small c_s'. The Mincerian specifications assume no explicit costs ($c_s = 0$) to schooling. The following Mincerian specification (5), remains the most widely used empirical earnings function –

$$In\, [Y(s,x)] = \alpha + \rho_s s + \beta_0 x + \beta_1 x^2 + \varepsilon \quad \text{.......} \quad (5)$$

Where, ρ_s, as above, is rate of return to schooling for all schooling levels; $Y(s,x)$, is the earnings at schooling level, s; x is work experience; and ε with $E(\varepsilon|s,x) = 0$, is a mean zero residual. The development of (5) is inspired by frameworks used by Mincer (1958, 1974). Mincer (1958) used the Compensating Differences Model; and Mincer (1974) used the Accounting – Identity Model. Although both frameworks are algebraically similar, they are conceptually, different, having contents that are economically different (Heckman et al. 2006). To give an understanding of the specifications of (5), I will further explore the Compensating Differences and Accounting – Identity Models used by Mincer (1958, 1974)

Following the accounts of Heckman et al. (2006). In explaining how a difference in levels of education determines wage differential over lifetimes, Jacob Mincer (Mincer 1958), used the principle of Compensating Differences. The compensating differential is defined by the net present value of earnings over a lifetime. This entails netting costs associated with different levels of investment in education from the resultant present value of earnings over a lifetime. Using some assumptions – a key assumption in this framework is that individuals only forego earnings while schooling and no direct costs to schooling are incurred. Hence, the framework only considers implicit or opportunity costs and no explicit or direct costs to schooling. Other useful assumptions include – individuals having identical abilities and opportunities; perfect credit markets; certainty in the economic environment and occupational attainment based on the amount of schooling required for each occupation, which means being in certain occupations requires longer schooling. A key idea of the framework is that a compensating wage differential is necessary for occupations with longer periods of schooling (or education). Following the works of Heckman et al. (2006), a mathematical representation of the model is thus: The PV of earnings associated with schooling level *s* is given by –

$$V\,(s) \;=\; Y(s)\int_s^{\mathcal{T}} e^{-rt}\,dt \;=\; \frac{Y(s)}{r}\,(e^{-rs} \;-\; e^{-r\mathcal{T}}) \;\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \;(6)$$

Here, $\mathcal{T}$ is the length of working life, assumed not to be dependent on s, r is the interest rate that is externally determined. With s level (years) of schooling, $Y(s)$ is the associated earnings (annual) assumed to be fixed over the lifetime, of the individual. From (6), equating lifetime streams of earnings to the respective (expected) levels of schooling and taking logs, results in (7) –

$$\ln\, Y(s) \;=\; \ln\, Y(0) + rs \;+\; \ln(\,(1 \;-\; e^{-r\mathcal{T}})/(1 \;-\; e^{-r(\mathcal{T}-\,s)}))\ldots\ldots\ldots\ldots\ldots\ldots\ldots \;(7)$$

It is important to note that with allocations (schooling) determined by demand conditions, the model requires individuals to be indifferent among schooling choices for equilibrium across heterogeneous schooling levels (Heckman et al., 2006). Model (7) suggests that people with relatively high education levels (s) attained, relatively high earnings are achieved, in the labour market. Relating (7) to the Mincerian specification (5). From the arguments earlier, and knowing (7) is at equilibrium, r, is the internal rate of return known to be the discount rate at the different levels of schooling, this gives the present value of corresponding lifetime earnings streams. From (7), with a large $\mathcal{T}$ (the length of working life), the ρ_s in (5) which represents the percentage increase in earnings over a lifetime, associated with a given level, s of schooling, must equate to r, which is the market interest rate or interest rate determined externally. Then when, $\rho_s = r$ – then (5) yields a ρ_s that equates to the r in (7), the education market is deemed to be at equilibrium. Therefore in (5), with $\rho_s < r$, further schooling investment decision may not create additional value, and the reverse, $\rho_s > r$, will mean further schooling decision will suggest value creation. The Compensating Difference model used by Mincer (1958) has some assumptions that are deemed unrealistic, an example is the assumption that individuals have identical abilities, and the works of Heckman et al. (2006) went on to show some assumptions that specify (7) which aid the derivation of (5) are not empirically plausible. This includes assumptions that the future (and earnings thereof) is certain; and the absence of non-pecuniary costs and benefits of schooling and working. To further understand the foundations of (5), I will explore some of the later works of Mincer (1974) that also underpin the derivation of (5) used extensively in this analysis.

Following the account of Heckman et al. (2006), I will further elaborate on the accounting-identity model to show how it contributed to the foundation of Mincerian wage equation. Beyond earnings from investments in formal schooling (or education) capital, the works of Mincer (1974) stressed lifecycle earnings dynamics – analysing potential and observed earnings from human capital investments that accrue from formal schooling and work experience. With some assumptions inspired by the Dynamic Human Capital Investment Model (Ben-Porath, 1967) and distinct from key assumptions in Mincer (1958), the works of Mincer (1974) suggest that there may be heterogeneity (or variation) in returns even across individuals with the same levels of schooling (or education), s, due to heterogeneity in human capital (formal schooling and work experience profiles) across individuals. Potential earnings, $\mathcal{P}_t$, at time, t, is expressed as thus:

$$\mathcal{P}_t \equiv \mathcal{P}_{t-1}(1 + k_{t-1}) \equiv \prod_{j=0}^{t-1}(1 + \rho_j k_j)\mathcal{P}_0 \ldots\ldots (8)$$

At age t, ρ_t represents the average return to training investments; and $\mathcal{C}_t$ represents the costs of investment in training, which is a fraction (k_t) of $\mathcal{P}_t$. Hence, $\mathcal{C}_t$ can be expressed as

$\mathcal{C}_t = \mathscr{k}_t \mathcal{P}_t$. From (8), with an assumption of a constant rate of return to post-school investment over ages which equals $\rho_{0.}$ The following (9) holds:

$$\ln \mathcal{P}_t \equiv \ln \mathcal{P}_0 + s \ln(1 + \rho_s) + \sum_{j=s}^{t-1} \ln(1 + \rho_0 \mathscr{k}_j) \approx \ln \mathcal{P}_0 + s\rho_s + \rho_0 \sum_{j=s}^{t-1} \mathscr{k}_j \ldots\ldots (9)$$

With years spent in full-time investment, $\mathscr{k}_t = 1$, used to express formal schooling assumed to take place at the beginning of life, yielding a constant rate of return, ρ_s, across all years of schooling (Heckman et al.,2006). The approximation, $\ln \mathcal{P}_0 + s\rho_s + \rho_0 \sum_{j=s}^{t-1} \mathscr{k}_j$, is obtained for minmised rate of return to post-school investment, $\rho_{0.}$, and a minimised rate of return to schooling, ρ_s.

Further discussions of the account of Heckman et al. (2006) –

As earlier mentioned, the Accounting-Identity model of Mincer (1974) was partly inspired by the Dynamic Human Capital Investment Model (Ben-Porath, 1967). With an assumption of a linearly declining rate of post-school investment which equates to the amount of work experience at a given age, t. Also, with an assumption that $\mathcal{T}$, the length of working life is independent of s, the level of schooling.

Under these assumptions, (9) is restated as thus:

$$\ln \mathcal{P}_{x+s} \approx \ln \mathcal{P}_0 + s\rho_s + (\rho_0 \kappa + \frac{\rho_0 \kappa}{2\mathcal{T}})x - \frac{\rho_0 \kappa}{2\mathcal{T}} x^2 \ldots\ldots (10)$$

In deriving (10) from (9), it is important to note that, $\mathscr{k}_{x+s} = \kappa\,(1 - \frac{x}{\mathcal{T}})$, given that, $x = t - s$. (10) is the basis of the Mincerian earning function which accentuates the observed earnings as the difference between potential earnings and human capital investment costs. The Mincerian Earnings Function (without an error term) is given thus:

$$In\, Y(s, x) \approx \ln \mathcal{P}_{x+s} - \kappa\,(1 - \frac{x}{\mathcal{T}}) = [\ln \mathcal{P}_0 - \kappa] + \rho_s s + (\rho_0 \kappa + \frac{\rho_0 \kappa}{2\mathcal{T}} + \frac{\kappa}{\mathcal{T}})x - \frac{\rho_0 \kappa}{2\mathcal{T}} x^2 \ldots\ldots (11)$$

Crucial to this analysis, is the ρ_s, which is an average rate of return across all levels of schooling investments. This specification (11) accentuates that the log of earnings has a linear relationship to levels of schooling; log earnings have both a linear and quadratic relationship to work experience.

Furthermore, from (11), Mincer derived a model that specifically allows for variation in ρ_s and κ across individuals, specifying an error term. Using a random coefficient model, as thus:

$$In\,Y(s_i, x_i) = \alpha_i + \rho_{si} s_i + \beta_{0i} x_i + \beta_{1i} {x_i}^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (12)$$

With, $\overline{\alpha} = E(\alpha_i)$; $\overline{\rho_s} = E(\rho_{si})$; $\overline{\beta_0} = E(\beta_{0i})$; and $\overline{\beta_1} = E(\beta_{1i})$

Generalising, (12) can be written as (14), where $\varepsilon = \left[(\alpha - \overline{\alpha}) + (\rho_s - \overline{\rho_s}) + (\beta_0 - \overline{\beta_0})x + (\beta_1 - \overline{\beta_1})x^2\right]$ ………………………………………………………………. (13)

$$In\,Y(s_i, x_i) = \overline{\alpha} + \overline{\rho_s} s + \overline{\beta_0} x + \overline{\beta_1} x^2 + \varepsilon \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (14)$$

With the mean of coefficients equal to coefficients in the more generalised form: $\overline{\alpha} = \alpha$; $\overline{\rho_s} = \rho_s$; $\overline{\beta_0} = \beta_0$; and $\overline{\beta_1} = \beta_1$. Results in the more generalised form of Mincerian Earnings Function (5) as described above and hence:

$$In\,[Y(s, x)] = \alpha + \rho_s s + \beta_0 x + \beta_1 x^2 + \varepsilon \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (15)$$

From (15) above, I create an augmented model –

$$In\,\mathcal{W}_i = \alpha + \rho_s s_i + \beta_0 x_i + \beta_1 x_i^2 + \delta \mathrm{X}i + \varepsilon_i \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (16)$$

Where,

ρ_s, which is an average rate of return across all levels of schooling investments.

s_i is the years of schooling (*years_educ* for the initial analysis)

$In\,\mathcal{W}_i$ is the natural logarithm of hourly wage, reported by the individual, i (*In_earnings_h_usd*)

x_i & x_i^2, are the linear and quadratic forms of (measures of) work experience (*tenure*).

ε_i is an error term.

$\mathrm{X}i$ is a vector for other observed exogenous predictor variables.

A2.2: Construction of the Big 5 (Personality Traits) –

Conscientiousness

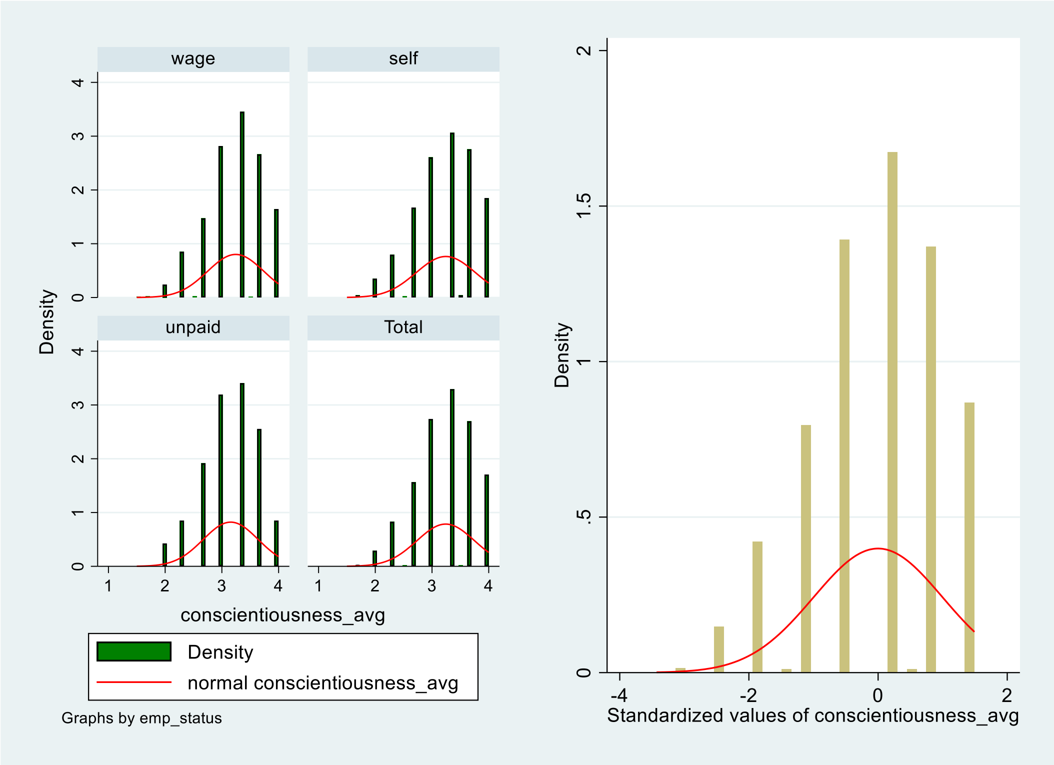


Figure A2.1: Trait of conscientiousness, disaggregated across employment categories; and standardised values for all employed.

Conscientiousness average (average of q02 q12 q17) – *conscientiousness_avg*. In module 6 of the STEP Household survey, the variable, *conscientiousness_avg* is used to capture propensity to follow socially prescribed norms. the STEP used a series of instruments to elicit this information from respondents (e.g., using a four-point frequency scale from that ranges from almost always to almost never, when doing a task are you very careful?). It is self-reported, or an indirect measurement as opposed to an assessed skill.

Trait or skill	*Questions in module 6 (G)*	*Items or survey instruments*
Conscientiousness Average – simple average of items.	Q.1.02	When doing a task, are you very careful?
	Q.1.12	Do you prefer relaxation more than hard work?
	Q.1.17	Do you work very well and quickly?

Table A2.1: survey instruments used to capture the trait, conscientiousness.

Where * is scaling reversal for negatively scored items. It is important to note that the *conscientiousness_avg* which is a simple average of all the instruments used within, it is a continuous variable that ranges from 1 – 4, with 1 meaning almost never which accentuates lowest evidence of trait; 4 almost always which accentuates high evidence of trait; 2 and 3 are medium levels of the trait.

Openness to Experience

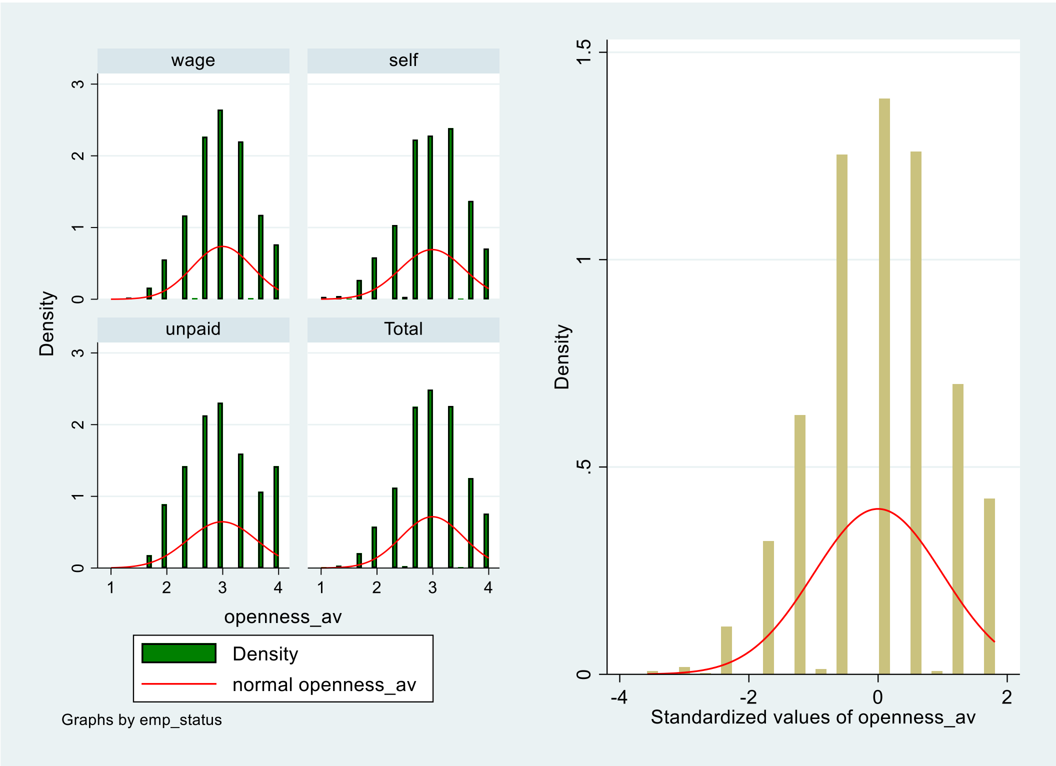


Figure A2.2: Trait of openness, disaggregated across employment categories; and standardised values for all employed.

Openness average overall score (average of q03 q11 q14) – *openness_av*.In module 6 of the STEP Household survey, the variable, *openness_av* is used to capture enjoyment of learning and new ideas. the STEP used a series of instruments to elicit this information from respondents (e.g., using a four-point frequency scale that ranges from *almost always* to *almost never*, do you come up with new ideas others haven't thought of earlier?). It is self-reported, or an indirect measurement, as opposed to assessed skill.

Trait or skill	***Questions in module 6 (G)***	***Items or survey instruments***
Openness Average – simple average of items.	Q.1.03	Do you come up with ideas other people haven't thought of before?
	Q.1.11	Are you very interested in learning new things?
	Q.1.14	Do you enjoy beautiful things, like nature, art, and music?

Table A2.2: survey instruments used to capture the trait of openness.

Where * is scaling reversal for negatively scored items. It is important to note that the *openness_av* which is a simple average of all the instruments used within, it is a continuous variable that ranges from 1 – 4, with 1 meaning *almost never* which accentuates the lowest evidence of trait; 4 almost always which accentuates high evidence of trait; 2 and 3 are medium levels of the trait.

Emotional Stability

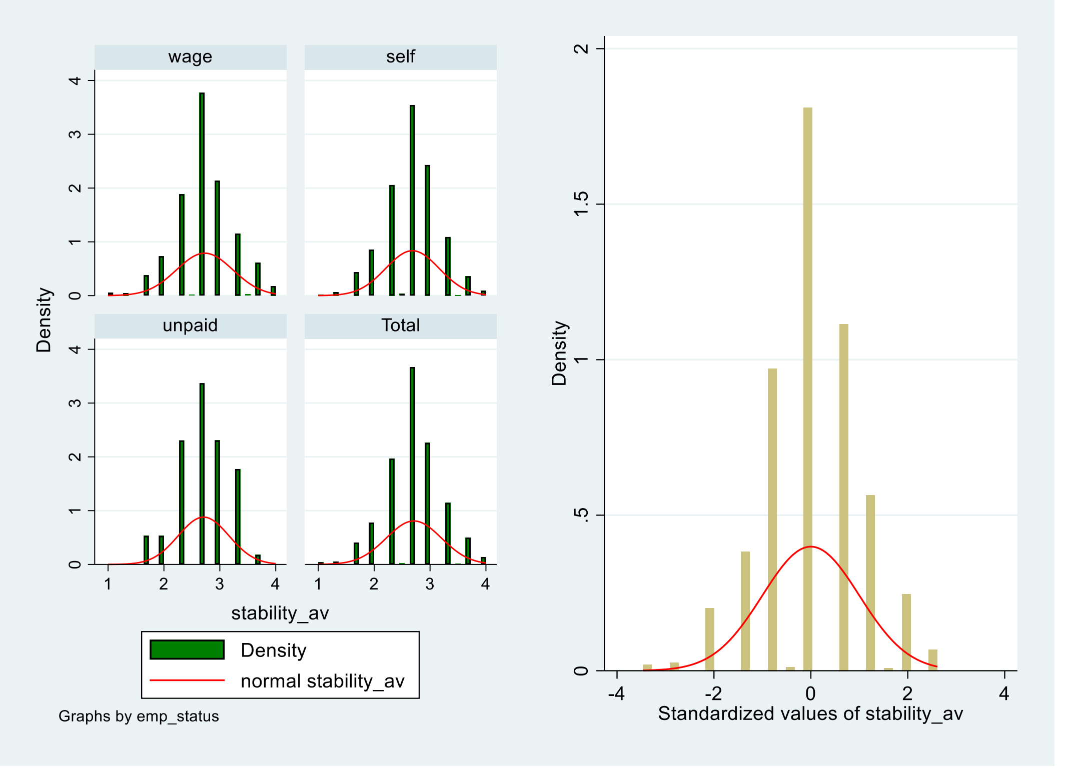


Figure A2.3: Trait of stability, disaggregated across employment categories; and standardised values for all employed.

Stability average overall score (average of q05 q10 q18) – *stability_av*. In module 6 of the STEP Household survey, the variable, *stability_av* captures the tendency to feel negative emotions. the STEP used a series of instruments to elicit this information from respondents (e.g., using a four-point frequency scale that ranges from *almost always* to *almost never*, do you worry a lot?). It is self-reported, or an indirect measurement, as opposed to an assessed skill.

Trait or skill	***Questions in module 6 (G)***	***Items or survey instruments***
Emotional Stability (Neuroticism) Average – simple average of items.	Q.1.05 *	Are you relaxed during stressful situations? *
	Q.1.10	Do you tend to worry?
	Q.1.18	Do you get nervous easily?

Table A2.3: survey instruments used to capture the trait of neuroticism.

Where * is scaling reversal for negatively scored items. It is important to note that the *stability_av* which is a simple average of all the instruments used within, it is a continuous variable that ranges from 1 – 4, with 1 meaning *almost never* which accentuates the *lowest evidence* of trait; 4 almost always which accentuates high evidence of trait; 2 and 3 are medium levels of the trait.

Agreeableness

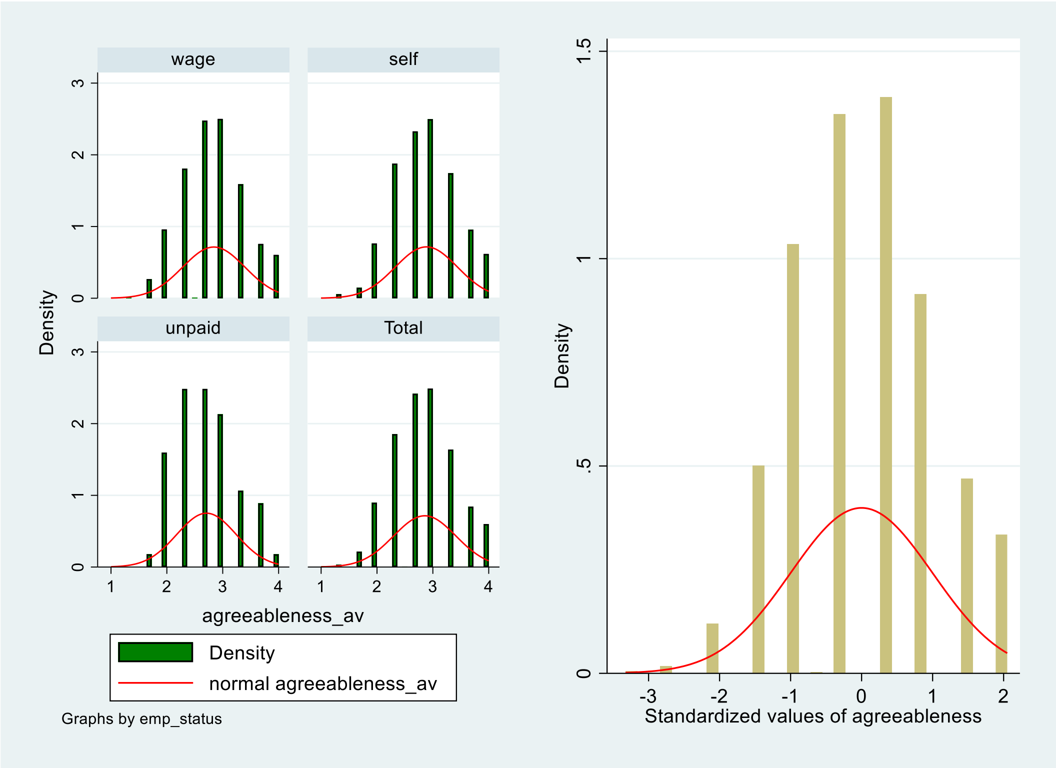


Figure A2.4: Trait of agreeableness, disaggregated across employment categories; and standardised values for all employed.

Agreeableness average overall score (average of q09 q16 q19) – *agreeableness_av.*

In module 6 of the STEP Household survey, the variable, *agreeableness_av* is used to capture cooperative orientation to others. The STEP used a series of instruments to elicit this information from respondents (e.g., using a four-point frequency scale from that ranges from almost always to almost never, do you forgive others easily?). It is self-reported, hence an indirect measurement as opposed to an assessed skill.

Trait or skill	***Questions in module 6 (G)***	***Items or survey instruments***
Agreeableness Average – a simple average of items.	Q.1.09	Do you forgive other people easily?
	Q.1.16	Are you very polite to other people?
	Q.1.19	Are you generous to other people with your time and money?

Table A2.4: survey instruments used to capture the trait of agreeableness.

Where * is scaling reversal for negatively scored items. It is important to note that the *agreeableness_av* which is a simple average of all the instruments used within, it is a continuous variable that ranges from 1 – 4, with 1 meaning *almost never* which accentuates the lowest evidence of trait; 4 almost always which accentuates high evidence of trait; 2 and 3 are medium levels of the trait.

Appendix A3: Tests for Instrument Validity —

IV Tests: Overidentification; Under-identification; Weak Instruments and Heteroskedasticity.

Outcome Variable	**zapvlit_c**
Dependent Variables	**years_educ**
	avg_yos_
Underidentification test (Anderson canon. corr. LM statistic):	18.145
Chi-sq(7) P-val =	0.006
Weak identification test (Cragg-Donald Wald F statistic):	2.601
Stock-Yogo weak ID test critical values: 5% maximal IV relative bias	16.880
10% maximal IV relative bias	9.920
20% maximal IV relative bias	6.160
30% maximal IV relative bias	4.760
10% maximal IV size	23.720
15% maximal IV size	13.340
20% maximal IV size	9.770
25% maximal IV size	7.910
Sargan statistic (overidentification test of all instruments):	3.440
Chi-sq(6) P-val =	0.6325
Endogeneity test of endogenous regressors:	11.806
Chi-sq(1) P-val =	0.003
IV heteroskedasticity test(s) using levels of IVs only	
Ho: Disturbance is homoscedastic	
Pagan-Hall general test statistic: 8.245 Chi-sq(7) P-val	0.607
Outcome Variable	**In_earnings**
Dependent Variable	**years_educ**
	avg_yos_
Underidentification test (Anderson canon. corr. LM statistic):	9.632
Chi-sq(7) P-val =	0.141
Weak identification test (Cragg-Donald Wald F statistic):	1.377
Stock-Yogo weak ID test critical values: 5% maximal IV relative bias	16.880
10% maximal IV relative bias	9.920
20% maximal IV relative bias	6.160
30% maximal IV relative bias	4.760
10% maximal IV size	23.720
15% maximal IV size	13.340

20% maximal IV size	9.770
25% maximal IV size	7.910
Sargan statistic (overidentification test of all instruments):	15.934
Chi-sq(6) P-val =	0.007
Endogeneity test of endogenous regressors:	11.119
Chi-sq(1) P-val =	0.003
IV heteroskedasticity test(s) using levels of IVs only	
Ho: Disturbance is homoscedastic	
Pagan-Hall general test statistic: 8.245 Chi-sq(7) P-val	0.075
Outcome Variable	**In_earnings**
Dependent Variable	**apvlit_c**
	avg_skill_
Underidentification test (Anderson canon. corr. LM statistic):	6.224
Chi-sq(7) P-val =	0.399
Weak identification test (Cragg-Donald Wald F statistic):	0.888
Stock-Yogo weak ID test critical values: 5% maximal IV relative bias	16.880
10% maximal IV relative bias	9.920
20% maximal IV relative bias	6.160
30% maximal IV relative bias	4.760
10% maximal IV size	23.720
15% maximal IV size	13.340
20% maximal IV size	9.770
25% maximal IV size	7.910
Sargan statistic (overidentification test of all instruments):	14.895
Chi-sq(6) P-val =	0.021
Endogeneity test of endogenous regressors:	15.557
Chi-sq(1) P-val =	0.008
IV heteroskedasticity test(s) using levels of IVs only	
Ho: Disturbance is homoscedastic	
Pagan-Hall general test statistic: 8.245 Chi-sq (7) P-val	0.390

Table A3.1: Tests of Instrument Validity, **taking both independent variables as endogenous (actual years of schooling & average years of schooling in the district; and individual level skill and average skills in the district)**. With standardised individual level skill (zapvlit_c); and log hourly earnings in USD (In_earnings_h_usd) as dependent variables. The reform dummy, quarter of birth, and interaction of the reform dummy and quarter of birth are instruments used in instrumenting the endogenous (independent) variables. This is the basis of the 2SLS-IV results presented. I have discussed this and the related 2SLS-IV outputs in the Conclusion (see Chapter 6).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	zapvlit_c				ln_earnings_h_usd				ln_earnings_h_usd			
yos	0.199***		0.274***	0.118***	-0.0620		-0.231	0.171***				
	(0.000)		(0.001)	(0.000)	(0.332)		(0.141)	(0.000)				
avg_yos		0.363***	-0.137*	0.148		-0.108	0.306*	-0.422**				
		(0.000)	(0.027)	(0.125)		(0.315)	(0.023)	(0.002)				
Skill									-0.00382		-0.00610	0.00991
									(0.356)		(0.233)	(0.052)
avg_skill										-0.0236	0.0141**	-0.0850
										(0.380)	(0.003)	(0.267)
_cons	-2.072***	-3.676***	-1.466***	-2.725**	1.338	1.764	0.0580	3.105*	1.350	4.801	-0.705*	13.75
	(0.000)	(0.000)	(0.000)	(0.002)	(0.063)	(0.114)	(0.903)	(0.014)	(0.078)	(0.310)	(0.012)	(0.272)
N	3156	3174	3156	3156	1832	1842	1832	1832	1842	1842	1842	1842
R-sq	0.216	.	0.022	0.214	.	.	.	.	.	.	.	.
adj. R-sq	0.216	.	0.022	0.213	.	.	.	.	.	.	.	.

Table A3.2: Presents similar specifications of 2SLS-IV outputs in Chapters 2 (with standardised reading proficiency as outcome) and 3 (with log hourly earnings in USD as output). However, here, rather than taking district-level and individual-level measures of human capital as endogenous (as in Chapters 2 and 3) columns (1) (3) (5) and (7) take individual-level schooling as endogenous; In (2) (4) (6) and (8) district level schooling is endogenous; In (9) and (11) individual skill is endogenous; and in (10) and (12) district-level skill is endogenous. Tests of the validity of Instruments are presented in Tables A3.3, A3.4, A3.5, A3.6. The reform dummy instruments the endogenous (independent) variable. This is the basis of the 2SLS-IV results presented here. I have discussed excerpts of these outputs and related outputs (presented in chapters 2, 3, and 4) in the Conclusion (Chapter 6). The discussions include outputs of the tests in the subsequent Tables (A3.3, A3.4, A3.5, A3.6).

Outcome Variable	**zapvlit_c**
Dependent Variables	**years_educ**
	avg_yos_
Underidentification test (Anderson canon. corr. LM statistic):	10.807
Chi-sq(7) P-val =	0.001
Weak identification test (Cragg-Donald Wald F statistic):	10.834
Stock-Yogo weak ID test critical values:	
10% maximal IV size	16.380
15% maximal IV size	8.960
20% maximal IV size	6.660
25% maximal IV size	5.530
Sargan statistic (overidentification test of all instruments):	0.000
Chi-sq(6) P-val = (equation exactly identified)	
Endogeneity test of endogenous regressors:	3.944
Chi-sq(1) P-val =	0.047
IV heteroskedasticity test(s) using levels of IVs only	
Ho: Disturbance is homoscedastic	
Pagan-Hall general test statistic: 3.399 Chi-sq(2) P-val	0.183
Outcome Variable	**In_earnings**
Dependent Variable	**years_educ**
	avg_yos_
Underidentification test (Anderson canon. corr. LM statistic):	**7.942**
Chi-sq(7) P-val =	0.005
Weak identification test (Cragg-Donald Wald F statistic):	7.964
Stock-Yogo weak ID test critical values:	
10% maximal IV size	16.380
15% maximal IV size	8.960
20% maximal IV size	6.660
25% maximal IV size	5.530
Sargan statistic (overidentification test of all instruments):	0.000
Chi-sq(6) P-val = (equation exactly identified)	
Endogeneity test of endogenous regressors:	13.480
Chi-sq(1) P-val =	0.0002
IV heteroskedasticity test(s) using levels of IVs only	
Ho: Disturbance is homoscedastic	
Pagan-Hall general test statistic: 6.247 Chi-sq(7) P-val	0.044

Outcome Variable	**In_earnings**
Dependent Variable	**apvlit_c**
	avg_skill_
Underidentification test (Anderson canon. corr. LM statistic):	**10.504**
Chi-sq(7) P-val =	0.001
Weak identification test (Cragg-Donald Wald F statistic):	10.547
Stock-Yogo weak ID test critical values:	
10% maximal IV size	16.380
15% maximal IV size	8.960
20% maximal IV size	6.660
25% maximal IV size	5.530
Sargan statistic (overidentification test of all instruments):	0.000
Chi-sq(6) P-val = (equation exactly identified)	
Endogeneity test of endogenous regressors:	6.140
Chi-sq(1) P-val =	0.013
IV heteroskedasticity test(s) using levels of IVs only	
Ho: Disturbance is homoscedastic	
Pagan-Hall general test statistic: 8.111 Chi-sq (7) P-val	0.017

Table A3.3: Tests of Instrument Validity, **taking actual years of schooling as endogenous and average years of schooling in the district as exogenous; and individual level skill as endogenous and average skills in the district as exogenous**. With standardised individual level skill (zapvlit_c); and log hourly earnings in USD (In_earnings_h_usd) as dependent variables. The reform dummy instruments the endogenous (independent) variable. This is the basis of the 2SLS-IV results presented here. I have discussed these outputs and related outputs in the subsequent Tables in the Conclusion (Chapter 6).

Outcome Variable	**zapvlit_c**
Dependent Variables	**years_educ**
	avg_yos_
Underidentification test (Anderson canon. corr. LM statistic):	40.467
Chi-sq(7) P-val =	0.000
Weak identification test (Cragg-Donald Wald F statistic):	40.953
Stock-Yogo weak ID test critical values:	
10% maximal IV size	16.380
15% maximal IV size	8.960
20% maximal IV size	6.660
25% maximal IV size	5.530
Sargan statistic (overidentification test of all instruments):	0.000
Chi-sq(6) P-val = (equation exactly identified)	
Endogeneity test of endogenous regressors:	3.944
Chi-sq(1) P-val =	0.047
IV heteroskedasticity test(s) using levels of IVs only	
Ho: Disturbance is homoscedastic	
Pagan-Hall general test statistic: 20.048 Chi-sq(2) P-val	0.000
Outcome Variable	**In_earnings**
Dependent Variable	**years_educ**
	avg_yos_
Underidentification test (Anderson canon. corr. LM statistic):	36.169
Chi-sq(7) P-val =	0.000
Weak identification test (Cragg-Donald Wald F statistic):	36.837
Stock-Yogo weak ID test critical values:	
10% maximal IV size	16.380
15% maximal IV size	8.960
20% maximal IV size	6.660
25% maximal IV size	5.530
Sargan statistic (overidentification test of all instruments):	0.000
Chi-sq(6) P-val =	
Endogeneity test of endogenous regressors:	13.480
Chi-sq(1) P-val =	0.0002
IV heteroskedasticity test(s) using levels of IVs only	
Ho: Disturbance is homoscedastic	
Pagan-Hall general test statistic: 9.772 Chi-sq(7) P-val	0.008

Outcome Variable	**In_earnings**
Dependent Variable	**apvlit_c**
	avg_skill_
Underidentification test (Anderson canon. corr. LM statistic):	**1.784**
Chi-sq(7) P-val =	0.182
Weak identification test (Cragg-Donald Wald F statistic):	1.783
Stock-Yogo weak ID test critical values:	
10% maximal IV size	16.380
15% maximal IV size	8.960
20% maximal IV size	6.660
25% maximal IV size	5.530
Sargan statistic (overidentification test of all instruments):	0.000
Chi-sq(6) P-val = (equation exactly identified)	
Endogeneity test of endogenous regressors:	6.140
Chi-sq(1) P-val =	0.013
IV heteroskedasticity test(s) using levels of IVs only	
Ho: Disturbance is homoscedastic	
Pagan-Hall general test statistic: 0.761 Chi-sq (7) P-val	0.684

Table A3.4: Tests of Instrument Validity, **taking actual years of schooling as exogenous and average years of schooling in the district as endogenous; and individual level skill as exogenous and average skills in the district as endogenous**. With standardised individual level skill (zapvlit_c); and log hourly earnings in USD (In_earnings_h_usd) as dependent variables. The reform dummy instruments the endogenous (independent) variable. This is the basis of the 2SLS-IV results presented here. I have discussed these outputs and related outputs in the subsequent Tables in the Conclusion (Chapter 6).

Outcome Variable	**zapvlit_c**
Dependent Variables	**years_educ**
Underidentification test (Anderson canon. corr. LM statistic):	28.568
Chi-sq(7) P-val =	0.000
Weak identification test (Cragg-Donald Wald F statistic):	28.811
Stock-Yogo weak ID test critical values:	
10% maximal IV size	16.380
15% maximal IV size	8.960
20% maximal IV size	6.660
25% maximal IV size	5.530
Sargan statistic (overidentification test of all instruments):	0.000
Chi-sq(6) P-val = (equation exactly identified)	
Endogeneity test of endogenous regressors:	2.568
Chi-sq(1) P-val =	0.109
IV heteroskedasticity test(s) using levels of IVs only	
Ho: Disturbance is homoscedastic	
Pagan-Hall general test statistic: 0.108 Chi-sq(2) P-val	0.742
Outcome Variable	**In_earnings**
Dependent Variable	**years_educ**
Underidentification test (Anderson canon. corr. LM statistic):	24.233
Chi-sq(7) P-val =	0.000
Weak identification test (Cragg-Donald Wald F statistic):	24.532
Stock-Yogo weak ID test critical values:	
10% maximal IV size	16.380
15% maximal IV size	8.960
20% maximal IV size	6.660
25% maximal IV size	5.530
Sargan statistic (overidentification test of all instruments):	0.000
Chi-sq(6) P-val =	
Endogeneity test of endogenous regressors:	12.556
Chi-sq(1) P-val =	0.0004
IV heteroskedasticity test(s) using levels of IVs only	
Ho: Disturbance is homoscedastic	
Pagan-Hall general test statistic: 5.318 Chi-sq(7) P-val	0.021

Outcome Variable	**In_earnings**
Dependent Variable	**apvlit_c**
Underidentification test (Anderson canon. corr. LM statistic):	**13.482**
Chi-sq(7) P-val =	0.0002
Weak identification test (Cragg-Donald Wald F statistic):	13.566
Stock-Yogo weak ID test critical values:	
10% maximal IV size	16.380
15% maximal IV size	8.960
20% maximal IV size	6.660
25% maximal IV size	5.530
Sargan statistic (overidentification test of all instruments):	0.000
Chi-sq(6) P-val = (equation exactly identified)	
Endogeneity test of endogenous regressors:	5.405
Chi-sq(1) P-val =	0.020
IV heteroskedasticity test(s) using levels of IVs only	
Ho: Disturbance is homoscedastic	
Pagan-Hall general test statistic: 6.457 Chi-sq (7) P-val	0.011

Table A3.5: Tests of Instrument Validity, **taking individual years of schooling; and skill as endogenous**. With standardised individual level skill (zapvlit_c); and log hourly earnings in USD (In_earnings_h_usd) as dependent variables. The reform dummy instruments the endogenous (independent) variable. This is the basis of the 2SLS-IV results presented here. I have discussed these outputs and related outputs in the subsequent Tables in the Conclusion (Chapter 6).

Outcome Variable	**zapvlit_c**
Dependent Variables	**avg_yos**
Underidentification test (Anderson canon. corr. LM statistic):	59.110
Chi-sq(7) P-val =	0.000
Weak identification test (Cragg-Donald Wald F statistic):	60.193
Stock-Yogo weak ID test critical values:	
10% maximal IV size	16.380
15% maximal IV size	8.960
20% maximal IV size	6.660
25% maximal IV size	5.530
Sargan statistic (overidentification test of all instruments):	0.000
Chi-sq(6) P-val = (equation exactly identified)	
Endogeneity test of endogenous regressors:	11.784
Chi-sq(1) P-val =	0.0006
IV heteroskedasticity test(s) using levels of IVs only	
Ho: Disturbance is homoscedastic	
Pagan-Hall general test statistic: 4.586 Chi-sq(2) P-val	0.032
Outcome Variable	**In_earnings**
Dependent Variable	**avg_yos**
Underidentification test (Anderson canon. corr. LM statistic):	52.748
Chi-sq(7) P-val =	0.000
Weak identification test (Cragg-Donald Wald F statistic):	54.244
Stock-Yogo weak ID test critical values:	
10% maximal IV size	16.380
15% maximal IV size	8.960
20% maximal IV size	6.660
25% maximal IV size	5.530
Sargan statistic (overidentification test of all instruments):	0.000
Chi-sq(6) P-val =	
Endogeneity test of endogenous regressors:	4.703
Chi-sq(1) P-val =	0.030
IV heteroskedasticity test(s) using levels of IVs only	
Ho: Disturbance is homoscedastic	
Pagan-Hall general test statistic: 8.328 Chi-sq(7) P-val	0.004

Outcome Variable	**In_earnings**
Dependent Variable	**Avg_skill**
Underidentification test (Anderson canon. corr. LM statistic):	**4.776**
Chi-sq(7) P-val =	0.029
Weak identification test (Cragg-Donald Wald F statistic):	4.783
Stock-Yogo weak ID test critical values:	
10% maximal IV size	16.380
15% maximal IV size	8.960
20% maximal IV size	6.660
25% maximal IV size	5.530
Sargan statistic (overidentification test of all instruments):	0.000
Chi-sq(6) P-val = (equation exactly identified)	
Endogeneity test of endogenous regressors:	2.061
Chi-sq(1) P-val =	0.1511
IV heteroskedasticity test(s) using levels of IVs only	
Ho: Disturbance is homoscedastic	
Pagan-Hall general test statistic: 6.198 Chi-sq (7) P-val	0.0128

Table A3.6: Tests of Instrument Validity, **taking district level schooling; and skill as endogenous**. With standardised individual level skill (zapvlit_c); and log hourly earnings in USD (In_earnings_h_usd) as dependent variables. The reform dummy instruments the endogenous (independent) variable. This is the basis of the 2SLS-IV results presented here. I have discussed these outputs and the related outputs in the subsequent Tables in the Conclusion (Chapter 6).

Table A8: IV Tests.

A3.1-Tests of Endogeneity of Individual and District-Level Schooling and Skill

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	years_educ	years_educ	ln_earnings	avg_yos	avg_yos_	ln_earnings	Apvlit	apvlit	ln_earnings	avg_skill	avg_skill_	ln_earnings	years_educ	years_educ	zapvlit	avg_yos_	avg_yos_	zapvlit_c
0.qob_2_#1.p1985_	1.734***	1.734***		1.274***	1.274***		11.430	11.430		1.417	1.417		1.734***	1.734***		1.274***	1.274***	
	(0.000)	(0.000)		(0.000)	(0.000)		(0.247)	(0.247)		(0.599)	(0.599)		(0.000)	(0.000)		(0.000)	(0.000)	
1.qob_2_#0.p1985_	-0.042	-0.042		0.151	0.151		-4.184	-4.184		-0.937	-0.937		-0.042	-0.0415		0.151	0.151	
	(0.938)	(0.938)		(0.458)	(0.458)		(0.720)	(0.720)		(0.768)	(0.768)		(0.938)	(0.938)		(0.458)	(0.458)	
1.qob_2_#1.p1985_	1.925***	1.925***		1.425***	1.425***		13.410	13.410		2.459	2.459		1.925***	1.925***		1.425***	1.425***	
	(0.000)	(0.000)		(0.000)	(0.000)		(0.251)	(0.251)		(0.440)	(0.440)		(0.000)	(0.000)		(0.000)	(0.000)	
0.qob_3_#1.p1985_	-0.586**	-0.586**		-0.140	-0.140		-9.057*	-9.057*		-1.047	-1.047		-0.586**	-0.586**		-0.140	-0.140	
	(0.004)	(0.004)		(0.069)	(0.069)		(0.041)	(0.041)		(0.386)	(0.386)		(0.004)	(0.004)		(0.069)	(0.069)	
1.qob_3_#0.p1985_	-0.085	-0.085		0.0363	0.036		-22.28	-22.28		-3.113	-3.113		-0.085	-0.0849		0.036	0.036	
	(0.888)	(0.888)		(0.874)	(0.874)		(0.091)	(0.091)		(0.387)	(0.387)		(0.888)	(0.888)		(0.874)	(0.874)	
0.qob_4_#1.p1985_	-0.267	-0.267		-0.295***	-0.295***		2.770	2.770		0.567	0.567		-0.267	-0.267		-0.295***	-0.295***	
	(0.203)	(0.203)		(0.000)	(0.000)		(0.544)	(0.544)		(0.649)	(0.649)		(0.203)	(0.203)		(0.000)	(0.000)	
1.qob_4_#0.p1985_	-0.040	-0.040		1.334***	1.334***		-38.17**	-38.17**		1.918	1.918		-0.040	-0.0396		1.334***	1.334***	
	(0.950)	(0.950)		(0.000)	(0.000)		(0.006)	(0.006)		(0.612)	(0.612)		(0.950)	(0.950)		(0.000)	(0.000)	
years_educ_act			-0.040												0.196***			
			(0.441)												(0.000)			
Residuals (e)			0.162**			0.251**			0.002			0.015			-0.0625			-0.019
			(0.002)			(0.003)			(0.364)			(0.524)			(0.085)			(0.768)
avg_yos_						-0.122												0.091
						(0.142)												(0.142)
apvlit_c									0.002									
									(0.489)									
avg_skill_												-0.006						
												(0.802)						
_cons	10.02***	10.02***	1.070	9.524***	9.524***	1.937*	180.1***	180.1***	0.317	174.1***	174.1***	1.681	10.02***	10.02***	-2.089***	9.524***	9.524***	-0.869
	(0.000)	(0.000)	(0.064)	(0.000)	(0.000)	(0.025)	(0.000)	(0.000)	(0.518)	(0.000)	(0.000)	(0.680)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.179)
N	3061	3061	1771	3078	3078	1780	3078	3078	1780	3078	3078	1780	3061	3061	3061	3078	3078	3078
R-sq	0.012	0.012	0.185	0.037	0.037	0.029	0.011	0.011	0.097	0.002	0.002	0.033	0.012	0.012	0.277	0.037	0.037	0.012
adj. R-sq	0.010	0.010	0.184	0.035	0.035	0.028	0.009	0.009	0.096	-0.000	-0.000	0.032	0.010	0.010	0.277	0.035	0.035	0.011

Table A3.7: Endogeneity Tests

Appendix A4: DiD Post-Estimations and Parallel Trend —

Post-estimation tests and non-parametric test for Parallel Trend:
Effects of Father's Post-Secondary Schooling (father_educ=1) on Education and Skill (reading proficiency)—Pre- and Post-Reform.

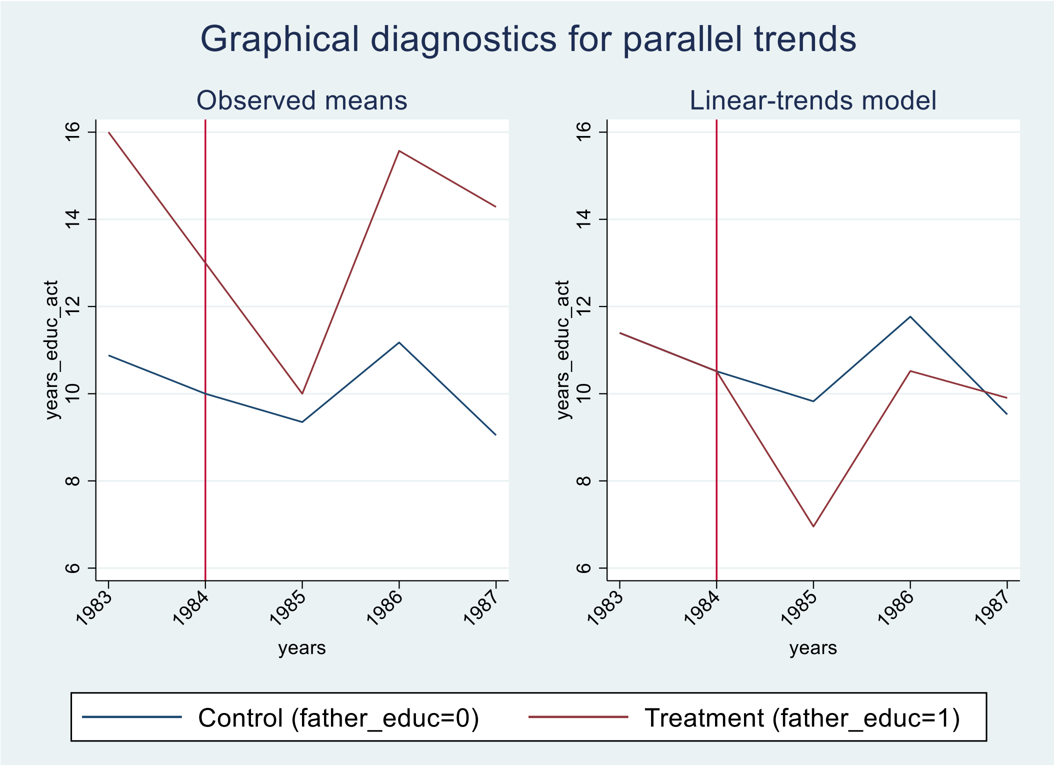


Figure A4.1: Effects of Father's Education on the Education of their ward.

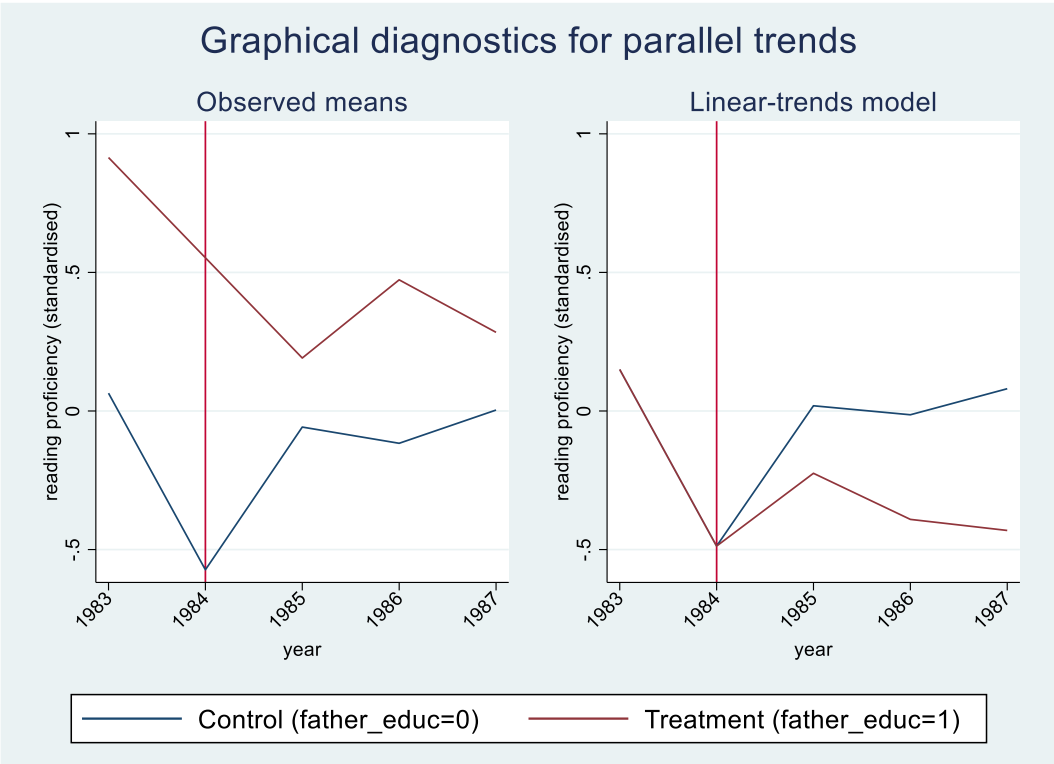


Figure A4.2: Effects of Father's Education on the Adult Skill (reading proficiency) of their ward.

Appendix A5: Data Appendix

Data Description, Specifications, Descriptive Evidence and Discussions

Introduction to the STEP Surveys and the place of skills.

The consideration for skills and personality traits as they relate to employability distinguishes this work from previous studies such as the works of Thias and Carnoy (1969); Johnson (1972); Knight and Sabot (1987) that focus on qualifications providing estimates returns to education in Kenya. In this work, I estimate returns to qualifications, skills, and personality traits as measures of human capital that accrue from schooling, in Kenya.

Beyond qualifications, the focus on skills and personality traits in this study is an attempt to give more understanding to recent arguments on the quality of schooling and the place of skills for growth in non-OECDs. The Skills Towards Employability and Productivity (STEP) surveys employed useful techniques and survey instruments to elicit data on qualifications[122] and a wide range of skills that exceed those acquired from formal education. Hence, the STEP surveys present measures (variables) that capture the quality[123] and what may be referred to as the quantity of (years spent in) formal education; and other measures that are useful proxies of human capital[124]. The STEP surveys collect a wide range of data on skills – this includes cognitive skills and non-cognitive[125] skills or personality traits of respondents.

The simultaneity of estimating returns to skills and schooling provides the added utility of evaluating the formal educational system, and the labour market in Kenya. It has been argued that cognitive skills are developed or nurtured through schooling (Barone and Van de Werfhorst, 2011) hence, cognitive skills may be considered school based. By school-based in their basic forms, I mean skills developed through numeracy and literacy[126] which are known to be the basis of productivity-enhancing skills required for information processing, critical thinking, and the ability to solve abstract problems that are rewarded in employment. Cognitive skills are the most useful of skills and they are known to be 'primary' skills – in

---

[122] Qualifications is taken to mean credentials and years of schooling.
[123] Quality is taken to mean skills from formal education.
[124] Here, human capital is taken to mean broad skills of individuals that may accrue from schooling and other (mainly work) experiences.
[125] I refer to non-cognitive skills as broad skills including job-related skills and personality traits that are not expressly referred to as cognitive skills.
[126] Literacy entails reading and writing.

the sense that all other known categories of skills including socio-emotional traits and job-related skills show some degree of dependence on the cognitive skills argued to mainly accrue from schooling (Barone and Van de Werfhorst, 2011). Hence, beyond the mere credentialing function of education systems, the school has a central role in skills acquisition and diffusion. This highlights the prominence and distinction of the formal education systems worldwide, hence, the focus on the Kenya educational system for growth.

The STEP surveys deploy indirect[127] and direct techniques to create useful measures of the cognitive skills of respondents. The indirect measures capture the levels, in terms of the intensity and frequency of the use of reading, writing and numeracy skills; and the direct measure involves the use of reading literacy assessments[128] to assess the cognitive skills of respondents. Also, using indirect techniques, the STEP surveys collected data on socio-emotional skills – personality or behavioural and attitudinal traits[129] of respondents, these traits elsewhere referred to as non-cognitive skills, are deemed crucial to the prosperity of the society. This category of skills has gained more prominence in recent literature as they are increasingly ascribed skills for growth, in today's world (Yang and Lester, 2016).

The next subsections of this chapter include identifying the data source – World Bank's STEP Surveys – this is the rationale for the use of the Surveys in this work; the next is the description of the dataset, which includes a brief account of the sampling and weighting procedure; also included in this subsection is a description of the STEP HS from which the analytical sample is derived, and I present a diagram – that show a brief description of all analytical subsamples used in this work; I include a description of variables and a table that summarises all variables used in this work. Lastly, I present a summary of the features of the STEP survey (and associated samples) that may impact results, proposing steps to mitigate possible adverse impacts on results and interpretations of results.

The STEP Surveys – Identifying Source and Rationale for the use of STEP Surveys.

Initiatives of the World Bank aimed at collecting and freely providing data for empirical research for low- and mid-income countries (non-OECDs) have the potential to significantly

[127] Indirect techniques refer to series of self-reporting styles of eliciting information from respondents.
[128] Assessments refer to carefully designed tests that are administered to respondents to capture their cognitive abilities.
[129] Traits and skills used interchangeably.

raise the capacity of research that relates to the non-OECDs. This makes it possible to have similar data as those available in high-income countries (OECDs).

An example of such initiatives of the World Bank is the STEP Measurement Program which took responsibility for providing Employer and Household Surveys. The STEP Surveys draw on similar surveys that were fielded in OECDs, with guidance on the construction of aggregated skills indicators like those of the Survey of Adult Skills (SAS) by the Program for the International Assessment of Adult Competencies (PIAAC) for OECDs. With a lineage to the PIAAC, the STEP Measurement Program's HS and ES for non-OECDs provide similar datasets (and variables) as those of the PIAAC's SAS for OECDs (Pierre et al., 2014). Highlighting similarities and differences between the STEP and PIAAC surveys provides an understanding that will aid the comparability of the output of this research with similar research outputs that used the PIAAC survey.

The first wave of SAS was fielded in 2011/12 in about twenty-four (24)[130] OECD countries and the first wave of STEP was fielded in fewer (7)[131] non-OECD countries at about the same period as those of SAS. A distinction between the STEP and SAS surveys is that the SAS has a household survey only, but the STEP implemented household and employer surveys.

Datasets of the STEP surveys are publicly available on the World Bank's central data catalogue. This study will use the datasets in the second wave – the Household Survey (HS) for Kenya implemented in 2013; and the third wave – the Employer Survey (ES) for Kenya implemented in 2016/2017. Both datasets are sourced freely from the World Bank's central data catalogue. The use of the HS and ES made it possible to access data on the supply and demand for skills, with the household survey giving insights on the supply of skills and the employers' survey which provides insights on the demand for skills.

Variables in the STEP HS present household-level and individual-level information of selected persons from participating households. Selected persons[132] are within the age interval: (15-64) years and some information elicited from the respondents include their educational attainment; family background; use of skills; history of skills acquisition; and

[130]Countries include Australia, Austria, Belgium (Flanders), Canada, Czech Republic, Denmark, Estonia, Finland, France, Germany, Ireland, Italy, Japan, Korea, Netherlands, Norway, Poland, Russian Federation, Slovak Republic, Spain, Sweden, United Kingdom (England and Northern Ireland) and United States.
[131]Countries include Columbia, Bolivia, China (Yunnan Province), Ukraine, Vietnam, Sri Lanka, and Lao PDR.
[132] Selected persons are elsewhere referred to as survey respondents.

work status among others. Before a detailed description of the variables, a useful question to ask is how the sampling and weighting procedure was operationalised. This gives some insights into the degree of randomness or the reliability of the sample in terms of the representativeness of the sample to the Kenya population, and ultimately, this accentuates the fit of the sample to the objectives of this study. I now describe the STEP sampling and weighting procedure and the main sample for Kenya.

The STEP Surveys – Describing the Sample and Weighting procedure.

An objective of this research is to compare research outputs to similar research work in non-OECDs and OECDs, but beyond the need to compare outputs, a description of the standardised implementation, sampling, and weighting procedure of the STEP Surveys present some insights on the reliability of the data collected and the relevance of the variables created, for this study.

Standardised Implementation

The account of Pierre, Puerta and Valerio of the World Bank, in their work in 2014, the *STEP skills measurement surveys: innovative tools for assessing skills* presents some details of the standardised implementation that further accentuates the relevance and reliability of the STEP dataset for this research.

Unlike the SAS, most participating countries of the STEP HS which include Kenya have an urban target population however, a few participating countries like the Lao People's Democratic Republic (PDR) and Sri Lanka included respondents in rural and urban areas. Issues surrounding the target population of the Kenya STEP HS raise some concerns over the sample selection procedure as it may not be representative of the working-age population of Kenya, a country deemed to be predominantly rural. In line with STEP standardised implementation among participating countries, the sampling strategy in Kenya was designed to achieve a target population representing at least ninety-five per cent of the urban population deemed to be working age – (15-64) years. The STEP sampling strategy is biased toward the urban population. The difference in the sampling strategies between the STEP and SAS that implemented what is deemed to be more random sampling, meant comparability of research outputs are inhibited. Hence, it may not be possible to fully compare the outputs of this work, to other research outputs that use the SAS. However, the outputs of this research may remain comparable with other research outputs that use the

STEP surveys. The bias to the urban population in the STEP sampling strategy meant, that care is required in interpreting findings. The results of this analysis relate to the urban population in Kenya, instead of the entire population in Kenya.

Sampling (and Weighting) Procedure

Understanding the sampling design (and weighting procedure) is useful in understanding the representativeness of the sample to the population, hence, following the account of the World Bank's publication: *STEP Survey Weighting Procedures Summary (Based on the World Bank Weight Requirement), Kenya, March 28, 2014*. I therefore present a summary of the sampling design (and weighting) procedure for the Kenya STEP Household survey. Beyond the sample representativeness, understanding the sampling and weighting procedure gives further insights into better interpreting the research outputs, using the STEP HS sample.

The target population consists of the working age (15-64 – inclusive) in urban areas only. These persons are non-institutionalised in private dwellings. At the time of data collection, the target population only include residents of Kenya except non-nationals working for international organisations and foreign diplomats. Further exclusions from the target population include the unstable regions of Kenya, including the war-marred regions; it excludes seniors in hospices and those in homes for seniors; it excludes those in college dormitories, workers' quarters, halfway homes, and others in similar categories of dwelling; it excludes Kenya nationals living outside Kenya at the time of the survey; it also excludes those classified as Itinerants based on the 2009 Population Census classification. Excluding certain groups of individuals may be necessary for useful sampling purposes, for example, excluding those in temporary dwellings (that may have their primary/principal residence elsewhere) will improve the sampling process and help to manage the administration burden on the survey team.

Following the sampling and weighting procedure adopted by the STEP Skills Measurement Program of the World Bank (please see details in the publication[133]). For this analysis, the main sample created and used in this study is deemed representative of the working-age population in urban Kenya as the sampling was restricted to only include the working age

[133] STEP Survey Weighting Procedures Summary (Based on the World Bank Weighting Requirement), Kenya, March 28, 2014.

population in urban areas of Kenya. This impacts the interpretation of results obtained using the sample.

The constituent of the final[134] sample in the sampling design includes 4242 eligible households (hence 4242 selected persons from each of the eligible households are expected in a case 100% response rate) drawn from 268 PSUs. The actual population is a total of 2 692 378 Households and 12 487 375 persons in Urban Kenya, this makes up 32% of the total population in Kenya (based on 2009 Population Census). With a response rate of 91.8% achieved in the Kenya STEP HS, the overall sample size of 3, 894 for Kenya makes the main sample from which other analytical subsamples of interest are drawn for this analysis.

Analytical Samples

With a response rate of about 92%, the sample (main) size of the STEP HS for Kenya is 3 894, this is deemed representative of the working age in urban Kenya. From the sampling (and the World Bank Weighting) procedure followed by the STEP, the final sample that is deemed representative of the population should reflect the labour force participation in Kenya, hence:

Labour market status: active and inactive categories – main sample, STEP HS for Kenya

Labour market status variable **(*lm_status*)**	**Frequency**	**Frequency (%)**	**Mean (age)**
The employed (analytical sample 1)	2 422	62.26	31.750
Unemployed	571	14.68	27.606
Inactive	897	23.06	24.708
Total	3 890	100	29.518

Source: Author's elaboration of the STEP Household Survey for Kenya. World Bank.

[134] Number of activated households.

In the table of variables, *lm_status* – labour market status, shows some statistics of the main sample of the STEP HS for Kenya from which the analytical sample of interest is derived. Of interest is the employed which is a subset of the labour force defined by the variable, *lm_status*. The *lm_status* shows details of the employed, the unemployed and the inactive in urban Kenya. The frequency of the employed is 2 422 (62.26%), with a mean age of 31.8 years; the unemployed has a frequency of 571 (14.68%) and a mean age of 27.6 years; and the inactive has a frequency of 897 (23.06%) and a mean age of 24.7 years. The variable, *lm_status* suggests that the labour force participation rate in urban Kenya is about 77%, which means only about 23% of the labour force is inactive in urban Kenya.

Since the objective of this work is to analyse returns to credentials, skills and personality traits in employment, the focus of this study is on the employed, I identify the following analytical sample from the main sample of the Kenya STEP HS: The main analytical sample for this study is defined by the variable, *emp_status*, this gives insights to the **employment status** of respondents in urban Kenya.

Employment status, showing subsets of the employed – Wage, Self and Unpaid in Kenya

Employment status*)* variable ***(emp_status)***	**Frequency**	**Frequency (%)**	**Mean (age)**
Wage–employed. **(analytical subsample 1.1)**	1 369	56.52	30.871
Self–employed. **(analytical subsample 1.2)**	986	40.71	32.967
Unpaid family worker	67	2.77	31.821
Total	2 422	100	31.750

Source: Author's elaboration of the STEP Household Survey for Kenya. World Bank.

Employment status, disaggregated by gender – STEP HS for Kenya

Source: Author's elaboration of the STEP Household Survey for Kenya. World Bank.

From the Table and Figure for the variable, *emp_status* – employment status, this shows some statistic of the main analytical sample for this study. These give further insights into urban Kenya, by accentuating some statistic of the employment categories relevant to this study.

Since I am interested in further analysis of the variations in returns across the wage- and self-employed, I therefore identify two subsets of the employed (analytical sample 1), subsets – the wage-employed category (analytical subsample 1.1); and the self-employed category (analytical subsample 1.2) – are of interest in this analysis. The table and figure show details of analytical sample 1 as it relates to the subsets – analytical subsamples 1.1 and 1.2. With 2 422 respondents employed having a mean age of 31.8 years. The wage-employed make 56.5% (1 369 respondents) of the employed, with a mean age of 30.9 years; the self-employed make 40.7% (986 respondents) of the employed, with a mean age of 33 years. The employment rate in urban Kenya is 62.3%, mapping the employment categories to the urban population in Kenya suggests that the categories of interest – the wage- and self-employed – make a total of 97.2% of the employed, and the unpaid family workers make only 2.77% of the employed in urban Kenya. Evidence also suggests that, on average, the self-employed are about two years older than the wage-employed in urban Kenya.

Rationale for disaggregating across the informal, gender and age groups in Kenya.

Informal

Amidst the ongoing arguments on the prevalence of the informal sector in most developing countries, some researchers have attempted to explain the prevalence of the sector as a labour market segment that is peculiar to the region, owing to its inherent characteristics. Others attempt to explain the prevalence of the informal sector with an understanding of the competitive market forces in the region. The heterogeneity (in terms of the varying sizes) of the informal sector among developing countries still makes existing arguments inconclusive. Gunther and Launov (2012) in exploring the case of an urban labour market of Cote d'Ivoire, explain the prevalence of the informal sector as attractive to some categories of workers and a last resort to other categories of workers. I argue that understanding how credentials, skills and personality traits explain returns for the formal and informal within varying categories of the employed, may give a new understanding of the current prevalence of the informal sector in Kenya and non-OECDs. I disaggregate the main analytical sample of this study across the formal and informal sectors to understand possible factors that drive the prevalence of the informal sector.

The following Table shows that over 75% of the employed in urban Kenya, are informal. The mean age of the informal workforce in Kenya is 31.6 years, slightly less than the formal workforce with a mean age of 32.1 years.

***informal*, showing the informal status of all employed (for the main analytical sample)**

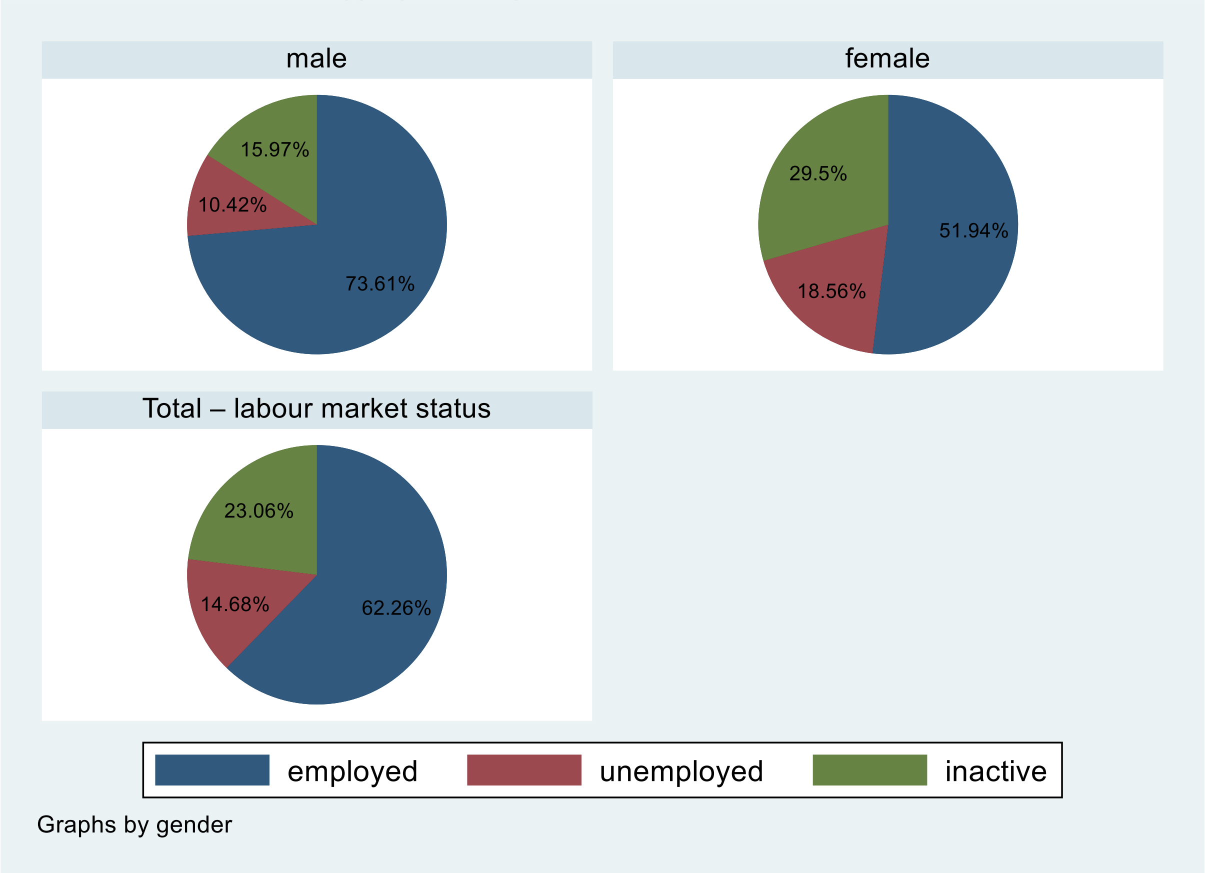
Informal variable (*informal*)	**Frequency**	**Frequency (%)**	**Mean (age)**
Formal	603	24.93	32.061
Informal	1 816	75.07	31.639
Total	2 419	100	31.744

Source: Author's elaboration of the STEP Household Survey for Kenya. World Bank.

With such a high prevalence of the sector – over seventy-five per cent (75%) of the employed in Kenya as informal – this compels a strong consideration for this category of workers in this analysis. I therefore disaggregate all analytical samples in this analysis to assess how credentials, skills, and personality traits impact returns to the informal, relative to the formal in the urban workforce in Kenya.

Gender

Labour market status, disaggregated by gender – main sample of STEP HS for Kenya.



Source: Author's elaboration of the STEP Household Survey for Kenya. World Bank.

From the above Figure for the variable, *lm_status* – labour market status, it is evident that the level of participation (labour force) of females differs from those of their male counterparts in all employment categories. The work of Anker (1998) of the International Labour Organisation puts forward some reasons to consider the specifics of the female gender, especially in survey data in empirical research in developing countries. This supports the need to analyse samples and see how the education and skills of the gender impact returns. The work of Vlasblom and Schippers (2005) in the OECDs suggests that the rising female labour force participation over the last decades is mainly driven by the 'changing times' (impacting the female labour supply), interpreted by changes (generational) in behaviours not explicitly related to their having children or educational attainment, but the fact that having a working wife and mother is the new normal in the OECD context. This is not to say educational attainment and fertility do not impact female labour force participation as evidence from their work reveals educational attainment and child-bearing impact labour force participation of females but deliberate efforts by government and related organisations aimed to encourage labour force participation of mothers and wives and also deliberate effort and behaviours (personality) of females to achieve the right timing of education, marriage and child-bearing have all contributed to the increase in female labour force participation in recent times. To give more understanding to the above arguments in non-OECD context, I therefore explore the impacts of personality traits, credentials, and skills on returns for the varying age groups of the female gender in non-OECDs, I operationalise this by disaggregating all analytical samples by sex.

Disaggregating the analytical sample by gender in urban Kenya suggests that females make up 43.68% of the employed. Within the employed, the females make 37.03% of the wage employed; 52.33% of the self-employed; and 52.24% of the unpaid family workforce. The males make up 56.32% of the employed, in urban Kenya. Within the employed, the males make about 62.97% of the wage employed; 47.67% of the self-employed; and 47.76% of the unpaid family workforce. Evidence from the analytical sample shows there are more males in wage employment and more females in self-employment. This is consistent with similar empirical works in non-OECDs.

Age groups

For all initial analyses in this work, I apply the age limit, 15-64 years, deemed to be the age range of most labour force around the world, this has been adopted by the STEP Skills Measurement Program and most empirical research in this area. However, the provisions of the 2007 Employment Act in Kenya, revised in 2012, stipulated that the minimum (legal) age for full-time employment in Kenya is 16 years of age, the Act also states conditions in which those in the 13-16 years age range can engage in what it described as *light* work. It is stated that children under 13 years of age are prohibited from any form of work. Also, the 2007 Employment Act in Kenya made a provision on the retirement age for public service (wage) employment in Kenya and this is set at sixty (60) years of age, at the maximum. Sequel to the age restrictions for full-time employment, in as much as I want to achieve comparability with similar research work in non-OECDs and OECDs by applying the usual (15-64) years limit to my analytical samples, I also aim to have findings that reflect the substance of everyday living in Kenya. I therefore identify age-varying analytical subsamples of the analytical sample. An example of this is the analysis of the public service wage employed, where I identify further analytical subsamples by limiting the age range to (16-60) years to reflect the substance of Kenya. I also complete analyses using the usual analytical samples – (15-64) years, I do the same for some other derived analytical subsamples, especially, increasing the lower limit from 15 years to 16 years, I also attach results to appendices, discussing material variances in result within commentaries.

Although descriptive evidence reveals certain age categories have lower levels of labour force participation, specifically, the (15-19) and the (45-64) age distribution, this is expected as their average returns to schooling are expected to reflect this, hence, interpreting results of these subcategories will consider this. Except otherwise stated, the (15-64) years age range is used in this work. I argue against limiting the analytical sample to (25-64) years or (30-60) years as other related research (that may have only considered credentials) has done, because most people complete their education at a certain age (Araki, 2020). This may not be appropriate for this research that aims to estimate returns to education in a country like Kenya, where school-leaving and school-entry ages differ significantly from those of the more stable OECDs that adopt the Key Stage system where students are placed in school year groups based on their age. It is not strange to find students significantly older or younger than their classmates, in the non-OECDs. Besides, this analysis considers all the employed not those requiring university-level qualifications. I therefore use credentials and years of schooling for all (15-64) years, by mapping the highest credentials obtained to expected years of schooling, I also include actual years of schooling as proxies of human capital, this will be discussed in subsequent subsections of this chapter. I disaggregate the analytical sample into five[135] different age groups to obtain useful descriptive evidence that helps analyse the impacts of credentials, skills, and personality traits on earnings and I interpret results accordingly.

I disaggregate the analytical sample 1 by the defined age groups. Disaggregating analytical sample 1 by age group reveals that the (25-34) years age group makes up about 44.51% of the employed and they dominate the wage-employed workforce, making up about 46.82% of the wage employed, they also make up about 43.10% of the self-employed but only about 17.91% of the unpaid family workforce. On the other end, the less dominating (15-19) age group only makes about 2.89% of the employed, they make about 3.43% of the wage-employed and about 1.52% of the self-employed. The (15-19) age group makes up about 11.94% of the unpaid family workforce. Relative to the rest of the age categories, the (15-19) age group is the most inactive, next to the (45-64) age group in Kenya, this is understandable as most within the (15-19) age category are expected to still be in full-time

[135] (15-19), (20-24), (25-34), (35-44) and (45-64).

education. The (expected) return of the (15-19) age category is expected to reflect insights into their levels of participation in urban Kenya.

The Variables

The World Bank's STEP Household Survey has household-level and individual-level information. These levels of information contain variables grouped in modules. An outline of the questionnaire, showing each module and its sub-components, is presented in Table 4 below. Within module 1 of the STEP dataset is the household-roster and dwelling-characteristic information that comprise variables at the household level; and a range of other modules that comprise information (variables) at the individual level. The individual-level modules present useful variables for this analysis, ranging from education & training; health; employment; self-reported cognitive and job-relevant skills; personality, behaviours, and preferences; language and family background; and reading literacy assessment. The objective of this subsection is to explore the STEP dataset. Hence, to describe the variables (and the related analytical samples) pertinent to this research.

Index for the STEP Household and Individual Questionnaire

	-Part A	Personality and Behaviour	44-45
	-Part B	Preferences	46-47
Module 7	Language and Family		48-50
8. INTERVIEWER IMPRESSIONS OF MODULES 2-7			51
Module 9	Reading Exercises		52
10. INTERVIEWER IMPRESSIONS OF MODULE 9 (READING EXERCISES)			53
11. RESULT CODES FOR BACK COVER			54
11. BACK COVER			

Source: World Bank (microdata/worldbank.org)

Variables within modules 2, 4, 5A, 6A, and 9 are the focus of this chapter. Specifically, modules 2, 5A, 6A, and 9 contain variables on educational attainment and skills, these include useful proxies to human capital; and module 4 contains variables on the employment of respondents, which is the basis of analytical sampling in this study. Module 4 contains variables on the earnings of the respondents. This is the main outcome variable used in this work.

Key Variables: measures of earnings, education, cognitive skills, and personality traits.

Within module 4 of the STEP dataset is the log hourly earnings in USD (*In_earnings_h_usd*), this is the main outcome variable used in this analysis. This is consistent with economic research estimating pecuniary returns to schooling. Module 2 of the STEP dataset presents variables on the educational attainment of the respondents, this includes credentials (*isced*) and years of education (*years_educ*) used in this analysis. In module 5, part A, the STEP dataset presents an indirect measurement of cognitive skills, this includes the frequency and intensity of the use of the skills of numeracy (*num*), reading (*read*), and writing (*write*); and module 9 presents direct measurements of cognitive skills through reading assessment test scores of the respondents. Within module 6 of the STEP dataset are aggregated measures of personality traits of the BFI – Big Five Inventory. This includes agreeableness_av, openness_av, conscientiousness_av, extraversion_av, and stability_av (emotional stability). These variables are central to this analysis.

Other Variables: work experience, participation, and school characteristics

Besides educational attainment, other useful measures of human capital include work experience. Within module 4 in the STEP dataset is the variable, tenure which captures months of work experience in a primary (or main) job. As the focus is on human capital that accrues from schooling, I include tenure as a control variable, and I do not attempt to construct potential experience[136] from dataset as some studies suggest. Besides the focus on human capital from schooling, the use of potential experience may not reflect the substance of Kenya, as school entrance and exit age for non-OECDs differ from those in OECDs.

Other predictor variables include participation in early childhood education, non-school training, government-recognised certification, and apprenticeships. Acknowledging the need to control for other known proxies of human capital that can significantly explain earnings, I have controlled for participation[137] as I am interested in returns to education attainment[138] and the skills. Other predictor variables accounted for include those typical of studies of this sort. These include age, gender, educational attainment of parents, and the number of economic shocks.

[136] Potential experience is constructed by ($A - Y_e - 6$), where A is the age of respondent, Y_e is the number of years of schooling and 6 is taken to mean the age when respondent commenced formal schooling.
[137] Participation in early childhood education, non-school training, government recognised certification and apprenticeships.
[138] Educational attainment are measures of human capital from formal education also referred to as educational capital in this work. This includes formal qualifications such as years of schooling and credentials.

Summary – Brief Description of some Variables

Table of Variable Description

Variable	Variable Type	Description – Survey Instrument.
Measures of Human Capital		
Isced	Discrete	Highest ISCED completed
years_educ	Continuous	Number of years of education corresponding to highest level completed
years_educ_act	Continuous	Actual years of education completed
Ece	Discrete	Attended pre-school
Apprenticeship	Discrete	Has completed an apprenticeship
Training	Discrete	Participated in a training course in the last 12 months
Certificate	Discrete	An industry-recognised or government certificate, not from a formal educational institution?
Measures of Cognitive Skills		
Num	Discrete	Numeracy overall score
Computer	Discrete	Frequency of computer use overall score
Read	Discrete	Length of material read overall score

Write	Discrete	Length of material written overall score
Measures of Personality Traits		
extraversion_av	Continuous	Extraversion (average of q01, q04 and q20)
conscientiousness_avg	Continuous	Conscientiousness (average q02, q12 and q17)
openness_av	Continuous	Openness (average of q03, q11 and q14)
agreeableness_av	Continuous	Agreeableness (average of q09, q16 and q19)
stability_av	Continuous	Emotional stability (average of q05, q10 and q18)
grit_av	Continuous	Grit (average of q06, q08 and q13)
Measures of Earnings		
earnings_h_usd	Continuous	USD Hourly labour earnings
ln_earnings_h_usd	Continuous	USD Log of hourly labour earnings
net_profit_usd	Continuous	USD Net monthly profit from business
Employment Categories/Samples		
lm_status	Discrete	Labour market status

emp_status	Discrete	Employment status
wage_worker	Discrete	Percent of wage workers
pub_emp	Discrete	Public or private sector employee
self_emp	Discrete	Self-employment dummy
business_size	Discrete	Size of business at the start
Informal	Discrete	Informal dummy
occtype_step	Discrete	Occupation type (STEP Aggregation)
Controls: Childhood and Background Characteristics		
Gender	Discrete	Gender
Age	Discrete	Age
Tenure	Continuous	Number of months in current jobs
Ses	Discrete	Socioeconomic status at age 15
Shocks	Discrete	Number of economic shocks before age 15
Chronic	Discrete	Chronic disease
father_mother (family intactness)	Discrete	Lived with mother and father at age 12
max_parent_educ	Discrete	Maximum of parents’ education
School Characteristics		

school_location	Discrete	School location
school_type	Discrete	Public school type attended

Source: Author's elaboration based on World Bank (microdata/worldbank.org)

General features of samples that may impact results – Summary.

Having identified and described the STEP datasets, the analytical samples, and the variables for this study, the dataset provided by STEP HS meets the needs of this study better than other available datasets. The STEP Skills Measurement Program for Kenya collected data on earnings, cognitive skills, non-cognitive skills or personality traits, and educational attainment of respondents aged (15-64) in urban Kenya. The datasets provided useful analytical samples and variables for this study but like most other empirical studies, the STEP datasets present some features that may adversely impact results. Drawing from publicly available documents published by the World Bank and from my descriptions within previous subsections of this chapter, I identify features of the datasets that may adversely impact results and implement safeguards to manage and effectively mitigate adverse impacts. For the analytical sample, I exclude all observations with missing entries.

A concern over the analytical samples identified is that the population from which the main sample of the STEP-HS is constructed is limited to urban areas only as all households involved in the sampling are in cities and not in rural Kenya. Given the inherent nature of the low- and mid-income countries, where access to rural households may be limited due to language, education, or issues around cultures, religions, and traditions. With this limitation, the data presented may significantly lack the required representation of the working-age population in countries with rural settlements, and this may result in some defects for country-level conclusions. Besides Lao PDR and Sri Lanka which included rural areas in their target population all other STEP countries only sampled their urban centres (based on their definitions of urban centres). However, from the sampling design, the target population should be representative of, at least ninety-five percent of the urban working-age population in all countries where the STEP Survey was implemented. The data weighting is managed by the Survey Methodologist who ensures consistency among the STEP participating countries, with regular adjustments (against benchmark variables such as age and gender) whenever new population counts are conducted. Unlike the STEP survey, the OECD's SAS has a target population of the resident adult population of the OECD countries. It is not limited to urban centres only as is the case of STEP surveys for non-OECDs. This is understandable as OECDs are high-income countries where there may be little or no limitations in data collection in areas deemed non-urban. A safeguard in place to mitigate the sampling issues requires communicating research output with some care. Hence, conclusions may relate to the urban working population instead of the working population in Kenya. As this also results in problems of comparability among other non-OECDs and OECDs that considered both rural and urban settlements in their sampling procedures, considerable care is taken to compare results among countries. Besides using the full analytical samples, I disaggregate all analytical samples and conduct separate analyses for the formal and informal to see how this impacts returns. Based on the scope of the survey and the literacy rates, the main sample sizes varied significantly among STEP participating countries, with the least observations (sample size) of 2 989 in Sri Lanka and the greatest of over 4000 observations (main sample size) in Macedonia. Response rates vary, with forty-three percent in Bolivia and ninety-eight percent in China (the Yunnan Province). These may result in discrepancies (e.g., from large variances between those with high observations/response rates and those with relatively low observations/response rates) that may accrue the measurement problems and issues of comparability of outputs. Hence, care is exercised in comparing outcomes for countries with (significantly) different characteristics – observations or sample sizes, response rates, and other important country-level characteristics. The other concerns over the samples are aspects typical of most surveys and not limited to STEP surveys, as most surveys (and variables) are self-reported. Current literature has recorded the effects of reticence in responses for certain types of self-reported survey instruments, especially for residents in the non-OECDs, particularly findings show that reticence is mainly found in the responses of survey instruments (questions) that relate to corruption. Serious biases caused by reticence in self-reported surveys, especially on survey instruments of socioeconomic significance (or wage/earnings related) may not be ruled out, completely, in non-OECD contexts. From the works of Karalashvili et al. (2018), using two alternative models of how reticence affects certain two-step survey questions, they developed a statistic that can reflect how much standard measures underestimate the proportion of all respondents who had a bribe interaction. This approach is explored in analysing earnings and returns for the public service wage-employed subsample. Crucial to this work are the skills variables. The main skills variables used in this work are known to be empirically validated. They are composites of several other variables (or survey instruments). Initial assessment of all variables used in the construction of the skills variables is necessary for understanding how instruments are administered and the responses achieved becomes imperative in understanding how the main skill variables are constructed – a combination (or composite) of multiple items to form a single indicator of the skill variable.

Variable Specification, Descriptive Evidence, and Discussions – Key Variables

The formal qualifications, cognitive skills, and personality traits (or behaviours) are known measures that may significantly explain earnings for the employed. This section of the chapter focuses on specific variables used as proxies of defined measures of human capital. I present *years of schooling* and *credential* variables for qualifications; *reading proficiency* variables for cognitive skills; and the Big Five Inventory (BFI) as variables for personality traits. I also present the USD log hourly earnings variable for returns to defined categories of the employed. Specifically, I describe these variables across analytical samples derived from the STEP dataset. Subsequent sections of this chapter include a detailed description of qualifications, skills, personality, and earnings variables across analytical samples. This includes descriptive evidence and specifications of variables, and I disaggregate, showing variations across informal status, age groups, and gender. As an extension to variable descriptions, I include arguments for the specifications, highlighting key literature, I then present a table that consists of the summary statistics of variables used in this work. Lastly, I highlight some possible strengths and weaknesses of variables as specified, citing their effectiveness (or ineffectiveness) as proxies to defined measures of human capital, this includes possible actions to mitigate weaknesses**.**

Descriptive Evidence, Empirical Specifications and Discussions – Human Capital.

Although human capital is taken to mean the entire stock of knowledge, experience, and skills including behaviours or personality traits possessed, there are several approaches and inexhaustive measures of human capital that remain controversial according to the work of Le, Gibson, and Oxley (2005). This study is concerned with specific measures of human capital that accrue or develop through education and training. Hence, the focus on formal qualifications, cognitive skills, and personality traits as measures of human capital may be influenced by initial education and training, which, in turn, may influence further education and training. In this study, I present and describe some variables as proxies of defined measures of human capital using the education-based approach.

Educational Attainment – as a measure of human capital in Kenya.

Commonly understood as the highest level of education an individual has completed, educational attainment is distinct from participation in education, training, certificates, and apprenticeships. Educational attainment entails defined knowledge and skills that succeed completion of certain years (key stages) of schooling or qualifications validated by forms of summative assessments[139] within the formal educational system. On the other hand, participation in education, training, certificate, and apprenticeships entail attendance (enrolment) of certain educational programmes, these programmes may be formal or informal (or non-statutory) as in the case of early childhood education in most countries (Roseth et al., 2016), participation may also be validated by demonstration of certain knowledge, skills, and experience. As earlier mentioned, educational attainment is the highest level[140] of formal education completed. In this analysis, I adopt credentials and years of schooling as variables for the highest level of formal education in my models. Harmon and Walker (1995) argued the place and prominence of *years of schooling* in determining outcomes of schooling, but the work of Dickson and Smith (2011), used data on the RoSLA– raising of the school leaving age in England and Wales. They found evidence that suggests, credentials (relative to years of schooling) impact returns to education as it is better able to explain employment in place of years of schooling used by Harmon and Walker (1995) and several other related empirical research. I therefore adopt both years of schooling, and credentials, as proxies and consider the differences in returns attributable to credentials/years of schooling in my discussion of results, if any. Within STEP HS, I use the ***years_educ_act*** and ***m2_q08b*** variables that define the actual number of years of education completed and Kenyan credentials obtained, respectively. For useful comparability of this research output to similar research works, I also use the variable, ***isced*** – highest ISCED[141] completed; and the variable, ***years_educ*** – number of years of education corresponding to the highest level (***isced***) completed. Within the STEP Household Survey, variables that capture educational attainment (proxies of qualification–credentials and years of schooling) as measures of human capital, in Kenya, are as follows:

[139] Summative assessments for measurement and evaluation (evidence) of defined knowledge and skills.
[140] Level of education is deemed to vary with qualifications and related skills that should accrue from it.
[141] International Standard Classification of Education (ISCED) presents a framework for presenting data (and qualifications) on formal schooling in a uniform and comparable manner. It facilitates the transformation of national education qualifications and data into internationally agreed categories which makes it possible to achieve cross-national comparisons.

Measures of educational attainment (formal qualifications) in Kenya

Variable	Survey Instrument – Question	Variable Type
isced	Highest ISCED (qualification) completed?	Discrete
years_educ	Number of years of education corresponding to highest level completed.	Continuous
years_educ_act	Actual number of years of education completed.	Continuous
m2_q08b	What is the highest level of formal education that you have completed?	Discrete

Source: own elaboration based on World Bank (microdata/worldbank.org)

Credentials

The ***isced*** reclassifies the credentials of respondents with Kenyan national credentials (***m2_q08b***) and provides the additional benefit of classifying the educational attainment of respondents with foreign qualifications to equivalent international credential categories (***isced***). While the ***isced*** is useful for comparability, the use of ***isced*** may come with some measurement bias introduced by its complex constructs, hence, to better reflect the substance of Kenya, I use the variable, ***m2_q08b*** along with the variable, ***isced***. I consider and discuss variances in the alternate use of these variables. I present results using ***isced*** and include results using the ***m2_q08b*** in the appendices. To understand how the Kenyan educational attainment classification is reclassified to the ISCED-97, an understanding of the Kenyan categories is useful. I therefore create specific categories of educational attainment within the ***m2_q08b*** variable, for this analysis.

***The m2_q08b* Variable – classification in Kenya –** the highest level of formal education that you have completed?

Within the STEP survey, the qualification (credential) of the respondents is recorded in the variable, *m2_q08b*. This is a response to the question: What is the highest level of formal education that you have completed? Hence, the *isced* variable is a modification or reclassification of the *m2_q08b* variable, following the ISCED-97. Details of the *m2_q08b* variable is within Module 2 (Education), question 8 of the STEP Household Survey.

Credential Classification in Kenya

Kenyan Categories – *m2_q08b* variable	**Educational Attainment in Kenya - Completed**
1	No level or grade completed (did not complete Primary)
2	Primary Education (completed Standards 1-8 – EACE/CPE/KCPE) – first 8 years of formal (primary) schooling – East African Certificate of Education/Certificate of Primary Education/Kenya Certificate of Primary Education
3	Post-Primary (completed Youth Polytechnic Craft Trade Certificate)
4	Secondary (completed Forms 1-6 - EACE, KJSE, KCE & KACE) – Kenya Junior Secondary Education; Kenya Certificate of Education; and Kenya Advanced Certificate of Education
5	Secondary (completed Forms 1-4 – EAACE & KCSE) – East Africa Advanced Certificate of Education; and Kenya Certificate of Secondary Education

6	TIVET Post-Secondary Certificate – Technical Industrial Vocational and Entrepreneurship Training, Certificate
7	TIVET Post-Secondary Diploma - Technical Industrial Vocational and Entrepreneurship Training, Diploma
8	TIVET National Polytechnic Higher Education –Technical Industrial Vocational and Entrepreneurship Training, Higher Education.
9	Primary Teacher Certificate
10	Teacher Training Diploma
11	Post-Secondary (not including Teacher Training) Certificate
12	Post-Secondary (not including Teacher Training) Diploma
13	University Degree (Bachelor)
14	Post-Graduate Diploma
15	Post-Graduate Degree (Masters)
16	Post-Graduate Degree (Doctorate/PhD)
97	Other

Source: The STEP Skills Measurement Survey for Kenya – Module 2, Question 8. World Bank.

ISCED – International Standard Classification of Education; and the variable, *isced* – highest ISCED completed?

Around the world, education systems vary in structure and content across countries, the International Standard Classification of Education (ISCED) presents a framework for presenting data (and qualifications) on formal schooling in a uniform and comparable manner. Specifically, the ISCED facilitates the transformation of national education qualifications (and data) into internationally agreed categories. This makes it possible to achieve cross-national comparisons. The ISCED was first developed by the United Nations Education Scientific and Cultural Organisation (UNESCO) in the 1970s and there have been several revisions afterward. The periodic updates are to reflect ongoing changes in education systems. The use of ISCED has gained general acceptance by researchers around the world, it is not just a reference classification within the United Nations Economic and Social Classifications, but it is widely accepted and used extensively in survey research and official statistics (Hoffmeyer-Zlotnik, 2008). Hoffmeyer-Zlotnik (2008) argues that using the ISCED-97 may result in miscalculations due to the complex constructs and combinations of the classification. This stresses the importance of a good knowledge of the ISCED in interpreting results, and the rationale for using the Kenya national qualification – years of schooling and credentials – I include these in appendices, highlighting material variances from the use of the *isced.* The latest version of the ISCED is the ISCED 2011 classification, a revision of the ISCED 1997, which provides a broader scope to aid more effective monitoring of global patterns in education. This provides more improved definitions of education systems around the world. Specifically, the 2011 update provided extensive revisions on sections that relate to tertiary and early childhood education. The UNESCO Institute for Statistics (UIS) is the custodian of ISCED and is responsible for the maintenance, new developments, revision, and updating of the ISCED. They provide guidance that supports a consistent (and effective) use of the ISCED for data collection and analysis to achieve international comparability. However, the STEP HS used the earlier version, the ISCED 1997 classification, as this was the version available at the time of the STEP survey.

ISCED-97 Classification and the UK Qualifications compared.

ISCED - 97 Levels/Sub-levels and Label.		Educational Attainment – Brief Description	Notes – Comparison to the UK.
0 – Pre-primary Education		This includes those with no formal schooling including those that did not complete primary education.	In the UK, this includes all in early childhood education and those who left primary school before age 11.
1 – Primary Education		Completion of primary schooling. Designed to provide basic education – skills of numeracy and literacy.	In the UK, this includes adult literacy and numeracy courses. This includes people who left secondary education at the age interval, 11-14.
2 – Lower Secondary Education	2A - 2B 2C	Covers a broad range of subjects. 2A gives access to 3A and 3B; 2B gives access to 3C; and 2C gives access to the labour market.	In the UK this includes those who left secondary education after age 14, without GCSEs.
3 – Upper Secondary Education	3A 3B 3C	More specialisation is offered – Sciences, Social-Sciences, and Arts. 3A gives access to 5A; 3B gives access to 5B; and 3C gives access to the labour market or other programmes at ISCED 3 or 4.	In the UK, ISCED 3C entails completion of GCSEs, Standard Grade, GNVQ/GSVQ Foundation & Intermediate, and NVQ Levels 1 & 2; ISCED 3A entails the completion of GCE A/AS, Higher/Advanced Higher Grade, GNVQ/GSVQ Advanced and NVQ Level 3.

4 – Post-secondary Non-tertiary Education	4A 4B 4C	This level straddles between levels 3 and 5 and typically lasts between 0.5 and 2 years. It gives access to either level 5 or the labour market. 4A gives access to 5A; 4B gives access to 5B; and 4C gives access to either the labour market or other ISCED 4 sub-levels.	In the UK, ISCED 4 is equivalent to Further Education training that gives access to Higher Education. This is mainly for those without a UK schooling background.
5 – First Stage Tertiary Education	5A 5B	5A is academic and theoretical and may be either medium/undergraduate – BA; or long/postgraduate taught – MA. 5B is more practical, technical, and specific to certain occupations – HNC & HND	In the UK, ISCED 5A attainment entails completion of BA, B.Sc., MA, and other postgraduate taught programmes that lead to M.Sc., PGCE, and PGDE. Attainment of ISCED 5B in the UK includes NVQ levels 4 & 5, CertHE and DipHE.
6 – Second Stage Tertiary Education		Leads to original research – adds new knowledge to a field.	Attainment of ISCED level 6 includes completion of a PhD.

Source: own elaboration based on UNESCO (2003)

The ISCED 0 – Pre-primary is designed to introduce children to the school environment. It is the preliminary stage of organised instruction. ISCED 1 – Primary is designed to give pupils sound literacy (reading and writing) and numeracy (basic mathematics). ISCED 2 – Lower-Secondary (2A, 2B & 2C) is a continuation from ISCED 1 but usually, more subject-focussed teaching from teachers that are more specialised than those in ISCED 1, in that, teachers at this level only teach a few subjects unlike teachers at ISCED 1 that are more of generalist subject teachers. Usually, at the end of 2C, students are deemed able to access and function in the labour market. ISCED 3 – Upper-Secondary (3A, 3B & 3C) is a continuation from ISCED 2, with more organised instructions along subject-matter lines with well-defined core subjects in place and options for students to specialise in either Sciences, Arts, or Business Studies. Duration at this level of studies can vary from 2 – 5 years. ISCED 4 – post-secondary (non-tertiary – 4A, 4B & 4C), typically, this lasts between 0.5 – 2 years of full-time equivalent studies, with much older students than those found at ISCED3. Programmes are not at the degree level, hence non-tertiary but straddles between studies at upper-secondary and post-secondary levels, not significantly advanced than ISCED3. This may contain some content taught at the tertiary level. ISCED 5 – First-stage tertiary (5A & 5B). This level is significantly more advanced than ISCED 4. It is known to be quite theoretical and leads to more advanced research degrees. Typically lasts a minimum of three years but this can increase to a maximum of seven years. With 5B deemed more technical/practical/occupational than 5A. ISCED 6 – Second-stage tertiary. This level involves original research that leads to the award of advanced research qualification like the PhD.

The variable, *isced* – highest ISCED completed; empirical specifications; and descriptive evidence.

The variable *isced* in the STEP data is one obtained using the survey instrument: Highest ISCED completed? Which elicits data on the highest qualification attained by respondents. Most respondents in Kenya respond to this by stating one of the categories of the m2_q08b variable, which is then reclassified into the appropriate ISCED-97 categories. The variable *isced*, summarised the ISCED-97 categories earlier described, this is presented within the STEP HS as thus:

Showing reclassification of the m2_q08b variable and summary of the ISCED categories

Value Assigned (categories)	**Educational Attainment (more verifications required)**	**ISCED 1997**
0	No formal schooling up to some years of Primary Education	None or < ISCED 1
1	Completion of Kenyan Primary Qualification	ISCED 1
2	Post-primary attainments – qualifications.	ISCED 2
3	Secondary and some post-secondary including TIVET.	ISCED 3 & 4A
4	Post-secondary Certificate and Diploma including Teacher Training – non-tertiary	ISCED 4B
5	Tertiary	ISCED 5 & 6

Source: The STEP Household Survey for Kenya. World Bank.

***isced* – variable specification for this study:**

To aid analysis of returns to credentials, using the *isced* variable, I innovate, creating dummy variables for distinct levels of the (*isced*) variable as thus: Respondents with no formal education and those with some formal schooling or some years of primary education (ISCED0) but without completion of the primary education (ISCED1) fall within the ISCED01 category. Hence, a derived binary variable, **ISCED01** is created. This is the reference category for this analysis. **ISCED1** is the completion of primary education**,** this makes survey respondents whose educational attainment or highest qualification is the completion of primary education. Respondents in this category are deemed to have developed basic cognitive skills – numeracy and literacy. **ISCED2** is the completion of lower secondary education. Respondents in this category in Kenya are those with basic education. The completion of this level of education shows that the foundation of lifelong learning is in place. The respondents have the option to join the labour market after completing this stage of education. The variable, **ISCED34A** elicits the educational attainment of respondents with upper-secondary and some post-secondary education. Respondents with some skilled or specialised education are captured by the derived binary variable, **ISCED4B** – Completion of this level means some studies under full tertiary education programmes and slightly above upper-secondary programmes are completed. Even though some programmes within this ISCED classification are deemed part of university foundation course, programmes within ISCED4 are not regarded as studies at tertiary education in Kenya. Although most in this category are vocational/technical. **ISCED56,** is a dummy variable that captures the educational attainment of all survey respondents with tertiary education. The *isced* (categories) – Descriptive evidence:

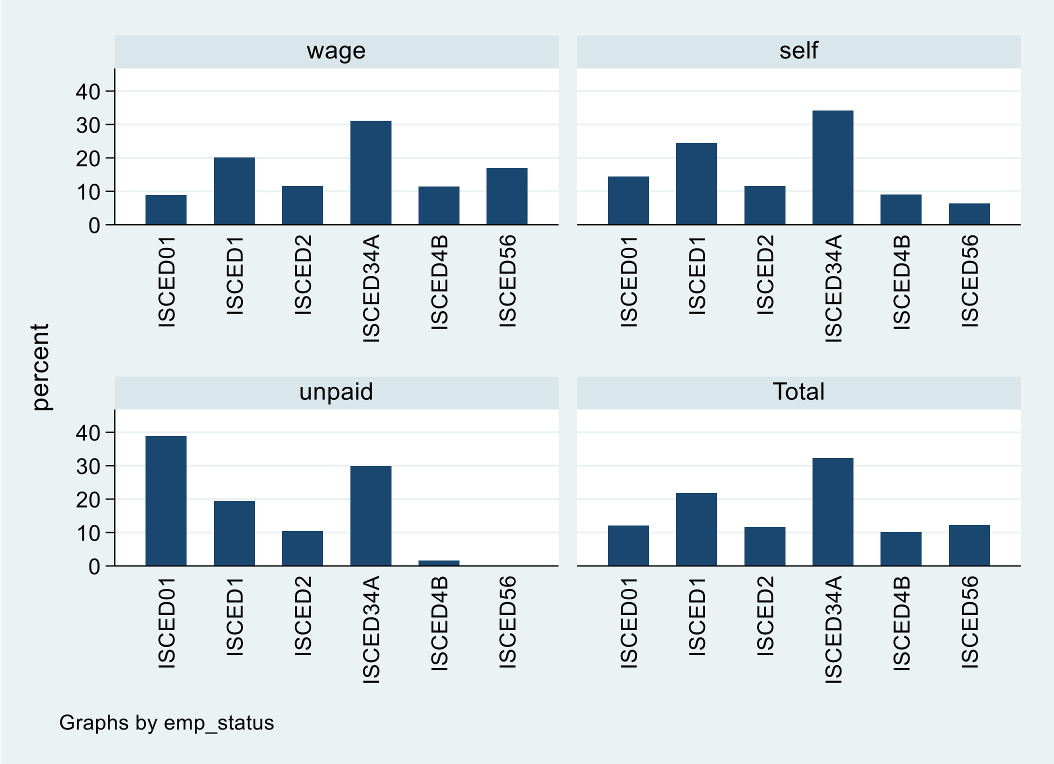
ISCED01 – 0 (reference category); ISCED1 – 1; ISCED2 – 2; ISCED34A – 3; ISCED4B – 4; and ISCED56 –5.

***isced* – frequency and mean age – main analytical sample**

	Wage	**Self**	**unpaid**	**Total**
Total (*isced*)				
Freq	1,361	984	67	2,412
% (Freq)	56.43	40.8	2.78	100
Age (mean)	30.892	32.978	31.821	31.769
Age (min)	16	15	18	15
Age (max)	64	64	60	64

Source: The STEP Household Survey for Kenya. World Bank.

***isced*– credential categories across all employment status – main analytical sample**



Source: Author's elaboration of the STEP Household Survey for Kenya. World Bank.

For each category of the *isced* within the analytical sample (all employed): 779 respondents out of 2412 employed are within the ISCED34A category of the *isced.* Qualifications within this category are upper-secondary, post-secondary but non-tertiary, this includes those that completed East Africa Advanced Certificate of Education (EAACE) or the Kenya Certificate of Secondary Education (KCSE) and those with a range of post-secondary qualifications in Kenya. This is the credential category for most, representing 32.3% of all the employed in Kenya. The mean age of workers within this credential category is 29.9 years, which is the youngest among the rest of the *isced* categories and about 2 years younger than the mean age of the employed which is 31.8 years. Comparing the wage and self-employed within the ISCED34A category of the *isced* show that 423 of the total 779 respondents within this category are wage-employed which makes 17.5% of the employed; and the self-employed within this category make about 13.9% of the employed, with a total of 336 of the 779 respondents in this credential category. Here, the self-employed within the ISCED34 category are 1.5 years older than the wage employed, this is consistent but slightly less than those of the entire sample, where the self-employed are over two years older than the wage employed. Following the ISCED34A category of *isced*, the next credential category with the highest number of respondents is the ISCED1 category, with a total of 527 respondents, representing about 21.9% of all employed in Kenya. These are respondents that completed primary[142] schooling, this may also include those with either some post-primary qualifications such as completion of the Youth Polytechnic Craft Trade Certificate and/or those with some years of lower-secondary schooling but without the completion of the lower-secondary qualification[143]. Respondents with qualifications in this credential category are relatively old, with a mean age of 32.3 years, which is 0.5 years older than the mean age of all employed which is only 31.8 years. Consistent with the rest of the credential categories, the wage employed dominates this category, although with a lower margin, relative to the self-employed. The wage employed with qualifications within this credential category make up 11.4% of the employed, which represents 274 of the 527 respondents within the ISCED1 credential category. Similarly, the self-employed within the ISCED1 credential category make up about 10% of the employed, with 240 respondents out of the 527 respondents in the ISCED1 category. Consistent with sample of all the employed, the self-employed are about 1.4 years older than the wage-employed within this ISCED1 category. Having discussed the credential categories of most respondents in the sample, understanding basic characteristics of respondents with the highest and lowest educational attainment will give some insights that will give more understanding to the econometric analysis in the analytic chapters. ISCED 56 is the *isced* credential category that accentuates some characteristics of respondents with the highest educational attainment in Kenya. All respondents within this credential category have tertiary qualifications. Specifically, for Kenya, this mainly represents those with either, some or all the university degrees. Degrees include bachelor's, master's doctorate; this may also include those with postgraduate diplomas. Of the 2414 respondents that are employed, only 294 respondents representing 12.2% of the employed have tertiary qualifications. The minimum age of respondents in this category is 21 which is consistent with required years of schooling for the educational attainment. The mean age of respondents in this credential category is 31.9 years, this is only slightly higher than the mean age of the employed (31.8 years) in Kenya. No member of the unpaid family workforce has qualifications in this credential category; 231 respondents out of the 294 respondents within ISCED56 category of the *isced* are in wage employment, with a mean age of 31 years, this makes up 9.6% of the employed; relative to wage employed, is the self-employed with only 63 respondents with qualifications within the ISCED56 credential category of the *isced*, with a mean age of 35 years. This makes up only 2.6% of the employed in urban Kenya.

[142] Completion of primary school in Kenya entails successful completion of first eight years of formal schooling, this leads to the award of the East Africa Certificate of Education (EACE) or the Certificate of Primary Education (CPE) or the Kenya Certificate of Primary Education (KCPE).

[143] Those that completed lower-secondary education are within the ISCED2 category of the *isced*.

Comparing the characteristics[144] of the wage- and self-employed, especially for those with the highest educational attainment – ISCED56 category of the *isced* – present interesting arguments that will be further discussed in the analytic chapters. The ISCED01, is the *isced* category for respondents with either no formal years of schooling or a few (under 8 years) years of formal schooling that did not achieve the ISCED1. It is important to note that respondents in this category may have no educational attainment (formal qualifications) but they may have participated in some certificate programmes, some training, apprenticeships or early childhood education which accrue to employment or job relevant skills. There are 289 respondents within the ISCED01 category of *isced* and this makes 12% of the employed in Kenya after the most employed categories (the ISCED34A and ISCED1) based on the evidence from analytical sample. The ISCED01 category have a mean age of 35 years, which is the highest among all the *isced* categories (not subcategories, as in the wage or self-employed within a category). Evidence from the sample suggests that the mean age of this category indicates rising educational attainment, which is consistent with the literature, as the mean age of all employed (analytical sample) is only 31.8 years, indicating a difference of 3.2 years. Interestingly, this is the only *isced* category with more self-employed than the wage employed. Within this category 121 (5.02% of the employed) respondents are wage employed with a mean age of 33.6 years; and 142 (5.89% of the employed) respondents are self-employed, with a mean age of 36.1 years. Considering the trends in the *isced* categories, specifically, how rising educational attainment impacts[145] employment status or increases the difference in labour market participation for the wage- and self-employed presents further argument for signalling in the labour market in Kenya. The wage employed with ISCED56 qualifications make up 9.6% of the employed in Kenya but the self-employed with similar qualifications only make up 2.6% of the employed. Comparing this to the ISCED01 category of the *isced,* where the wage employed only make 5.02% of the employed, with the self-employed making 5.89% of the employed. The trend suggests a widening that may be explained by employment choice and/or signalling from credentials. Evidence from simple descriptive analysis suggests that the significant difference in the mean age for wage- and self-employed, which is further accentuated by rising educational attainment appear to be characteristics that can explain employment and labour force participation in Kenya. Specifically, the difference in the mean ages of the self-employed and wage employed for the ISCED01 and ISCED56 are 2.5 (36.1-33.6) years and 4.02 (35.02-31) years respectively.

[144] Characteristics that relate and impact labour force participation (rates) and age.

These present further insights to be explored, this includes the nature[146] of self-employment at varying credential categories. Tentatively, relative to the wage-employed, the higher mean age of the self-employed suggests they are more experienced[147] and the higher difference in age for self-employed at higher educational attainment (or credential) categories suggests, the self-employment at higher ISCED categories may be more of entrepreneurships in lieu of lone-employment, as entrepreneurships may require useful skills, knowledge, experience, qualifications, and earnings which should accrue (over time).

Years of schooling

The variable, *years_educ* – years of education corresponding to highest level (*isced*) completed; *years_educ_act* – actual number of years of education completed.

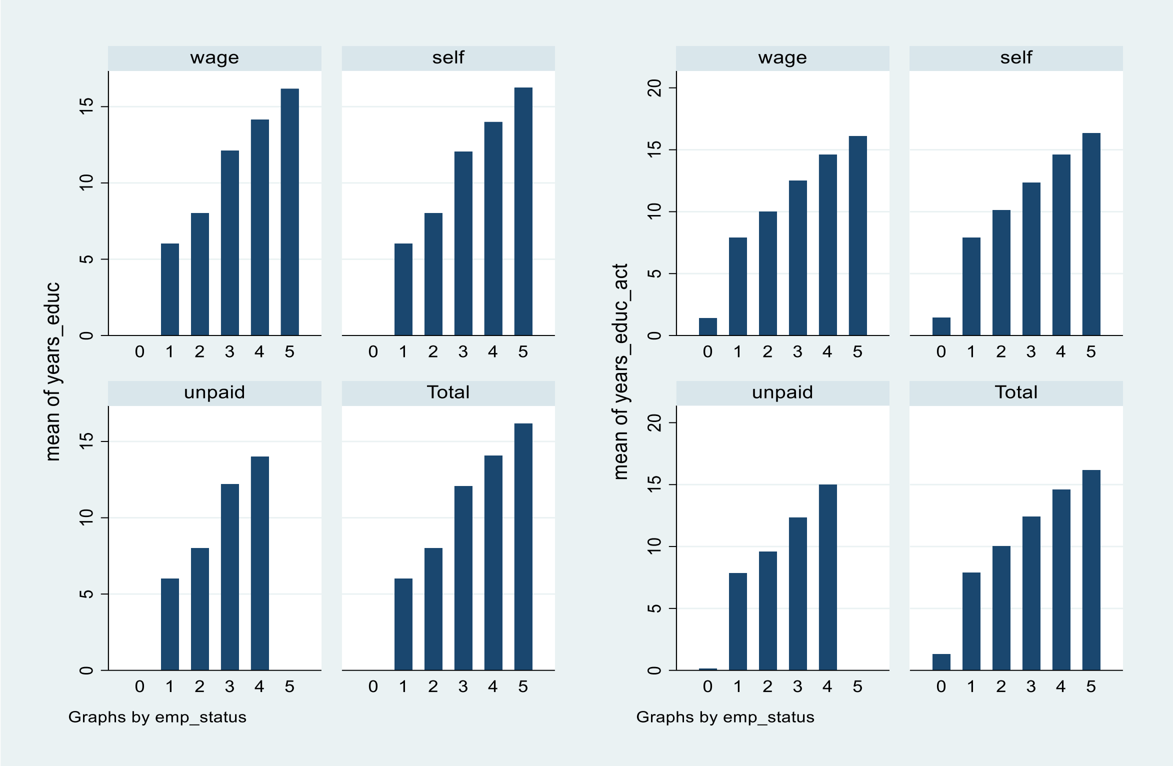
For years of education, I use the variable, *years_educ* – number of years of education corresponding to highest level (*isced*) completed. The *years_educ* maps the *isced* variable, hence, reflects the ***isced***, in continuous terms. The *years_educ* variable has additional (from *isced*) measurement defects, for example, ISCED0, translates to 0 *years_educ* which is not completely true, as the ISCED0 category of the *isced* may include respondents with some years of formal schooling as seen, those in ISCED0 have some primary education, with a mean 1.62987 *years_educ_act*. This makes an argument for the use of a variable that captures the actual years of schooling in Kenya. The variable, *years_educ_act* – actual years of education – this is useful and provide additional benefit that not only overcome the defects of *years_educ* as it relates to the *isced* but the *years_educ_act* better reflect the substance of Kenya. The *years_educ_act* is most useful in this context, as expected number of years of schooling (based on credentials) may differ significantly from actual years of schooling where students are made to repeat years (or levels) where they perform below a set threshold as is the case of Kenya and most non-OECDs that do not use the key stage system mainly used by most OECDs. For consistency, comparability and to reflect the substance in Kenya, I adopt the alternate use of both *years_educ* and *years_educ_act* variables. I include results of the *years_educ_act* variable in appendices, discussing material variances between the use of both variables. I report full results using the *years_educ* variable for consistency in my

[146] Here, nature of self-employment entails entrepreneurship (self-employment with some paid employees) or self-employment (lone).

[147] Years of experience may be crucial for a thriving business.

use of the *isced* for credentials. ISCED01 – 0 (reference category); ISCED1 – 1; ISCED2 – 2; ISCED34A – 3; ISCED4B – 4; and ISCED56 –5.

***years_educ* and *years_educ_act* compared – based on isced – sample (all employed)**



Source: Author's elaboration of the STEP Household Survey for Kenya. World Bank.

Descriptive evidence of years_educ and years_educ_act – sample of the employed

Percentiles		**Smallest**	***years_educ***		**Percentile**		**Smallest**	***years_educ_act***	
1%	0	0			1%	0	0		
5%	0	0			5%	0	0		
10%	0	0	Obs	2 404	10%	4	0	Obs	2 412
25%	6	0	Sum of wgt	2 404	25%	8	0	Sum of wgt	2 412
50%	12		Mean	9.516	50%	12		Mean	10.505
		Largest	Std. dev.	4.871			Largest	Std. dev.	4.479
75%	12	22			75%	14	20		
90%	16	22	Variance	23.726	90%	16	20	Variance	20.060
95%	16	22	Skewness	-0.496	95%	16	21	Skewness	-0.835
99%	18	22	Kurtosis	2.485	99%	18	22	Kurtosis	3.375

Source: Author's elaboration of the STEP Household Survey for Kenya. World Bank

The median (50th percentile) years of schooling for the employed in urban Kenya is 12, this strongly indicates that half of the respondents have less than twelve years of schooling and half have more than twelve years of schooling. With relatively (compared to median) low mean (average) years of schooling (9.5 years_educ and 10.5 years_educ_act), this indicates that most people in Kenya have more than the average years of schooling. Evidence from the Table 11 also indicates people spend an additional year of schooling, relative to the expected years of schooling based on their credentials (mean years_educ_act – mean years_educ = 0.988 years of schooling). This difference in years of schooling may not necessarily be the reality in Kenya (even if it is reasonable to say this may be true due to student repeating classes or spending more years by taking a gap year) but this may be a measurement problem resulting from the specifications of the *isced* variable.

Differences across years_educ and years_educ_act further accentuate material measurement biases that may result from the use years_educ, starting with the obvious, beyond the variance of the overall number observations (2404 years_educ and 2412 years_educ_act), the number of observations captured by each level (years of schooling) differ significantly, e.g., at 0 years of schooling, the year_educ variable captures 289 respondents and the years_educ_act variable only captures 209 respondents. The difference (about 80 respondents, which obviously have more years of schooling, up to 6 years of schooling) representing over 3% of the sample are wrongly specified by the years_educ and isced classifications. Another obvious misspecification, the years_educ (based on the isced) variable captured 8 respondents as having 22 years of schooling, relative to years_educ variable that only captured only 1 respondent as having 22 years of schooling. To reflect the substance of Kenya, the years_educ_act variable is considered in this analysis.

With a relatively high variability in the years_educ variable, the years_educ variable has a standard deviation of 4.870882 (compared to 4.478853 for the years_educ_act) years of schooling. Although both variables, years_educ and years_educ_act have the same range, (22 – 0) years of schooling and the same interquartile range (75th percentile – 25th percentile) of 6 years of schooling, the years_educ variable indicates the central half of the sample have between 6 and 12 years of schooling which further indicate that the quarter with the least years of schooling has under 6 years of schooling and the quarter of the sample with most years of schooling have more than 12 years of schooling. However, the years_educ_act variable indicates that the central half of the sample have between 8 and 14 years of schooling, which further indicate the quarter with the least educational attainment has under 8 years of schooling and the quarter with most educational attainment has over 14 years of schooling.

Descriptive evidence (years_educ; and years_edu_act) disaggregated by employment status

isced ***(years_educ)***	**wage**	**Self**	**unpaid**	**total**	***Isced*** ***(years_educ_act)***	**wage**	**self**	**unpaid**	**total**
Mean	10.225	8.787	5.851	9.516	Mean	11.153	9.881	6.493	10.505
Median	12	8	6	12	Median	12	11	8	12
SD	4.761	4.779	5.238	4.871	SD	4.293	4.436	5.395	4.479
Min	0	0	0	0	Min	0	0	0	0
Max	22	22	16	22	Max	20	22	15	22

Source: Author's elaboration of the STEP Household Survey for Kenya. World Bank

The Table above shows detailed breakdown of the descriptive evidence across all *isced* categories based on *years_educ* and *years_educ_act*; and across all employment categories and a total displaying aggregates for the main analytical sample (all employed). The Table above disaggregates the summary statistics across employment status, and this gives more insights to measurement issues that may arise from the use of the (years of schooling) variables described so far. The maximum years of schooling for the wage- and self-employed is 22 *years_educ*; but 20 and 22 *years_educ_act* for the wage- and self-employed, respectively. The suggest that *years_educ* variable overstates years of schooling for some categories of the *isced* (in this case for those in in ISCED56) and as previously discussed. the *years_educ* variable also understates the number of years of schooling for some categories of the *isced* (in this case for the ISCED01 category), this further support the need for the use of the *years_educ_act* variable as a substantive measure of years of schooling for the employed in Kenya. Table 12 further reveals breakdown (of key statistics) across employment categories. With a mean of 9.516223 *years_educ* for the employed, the mean *years_educ* for the wage (10.2249) and self-employed (8.786952) varies significantly with a difference of 1.44 *years_educ*; the *years_educ_act* shows some consistency (but significantly different) with this, with a mean of 10.50456 *years_educ_act* for the entire sample and 11.153 and 9.881098 *years_educ_act* for the wage and self-employed respectively, showing an actual difference in years of schooling of only 1.27 *years_educ_act* less than the difference in years of schooling captured by the *years_educ* variable. Although *years_educ* and *years_educ_act* variables show some consistency in median years of schooling for the main analytical sample and the wage employed, with both variables indicating half of all the employed and the wage employed have over 12 years of schooling and the other half have under 12 years of schooling with respect to both the *years_educ_act* and *year_educ* variables but both variables captured significantly different median years of schooling for the self-employed. The median number of years of schooling is 8 *years_educ* and 11 *years_educ_act* for the self-employed. The three years (years of schooling) difference captured by both variables further accentuate the usefulness of *years_educ_act* variable, that indicates that half of the self-employed have over 11 (in lieu of 8 *years_educ*) years of schooling and the other half have under 11 (in lieu of 8 *years_educ*) years of schooling. However, from descriptive evidence discussed so far, it is suggestive that the wage employed, on the average, have more years of schooling relative to the self-employed but evidence also suggest that the employment category with respondents having the greatest years of schooling, is the self-employed.

Disaggregating the educational attainment of the employed across gender, informal and age_group.

Gender

Disaggregating analytical sample (all employed) across gender reveals females make 43.62% (with males making up 56.38%) of the employed in urban Kenya; on the average, the females are about eight months younger than the males, this may be because of the difference in the mean years of schooling as the males have over 9 months (on the average) of schooling relative to the females.

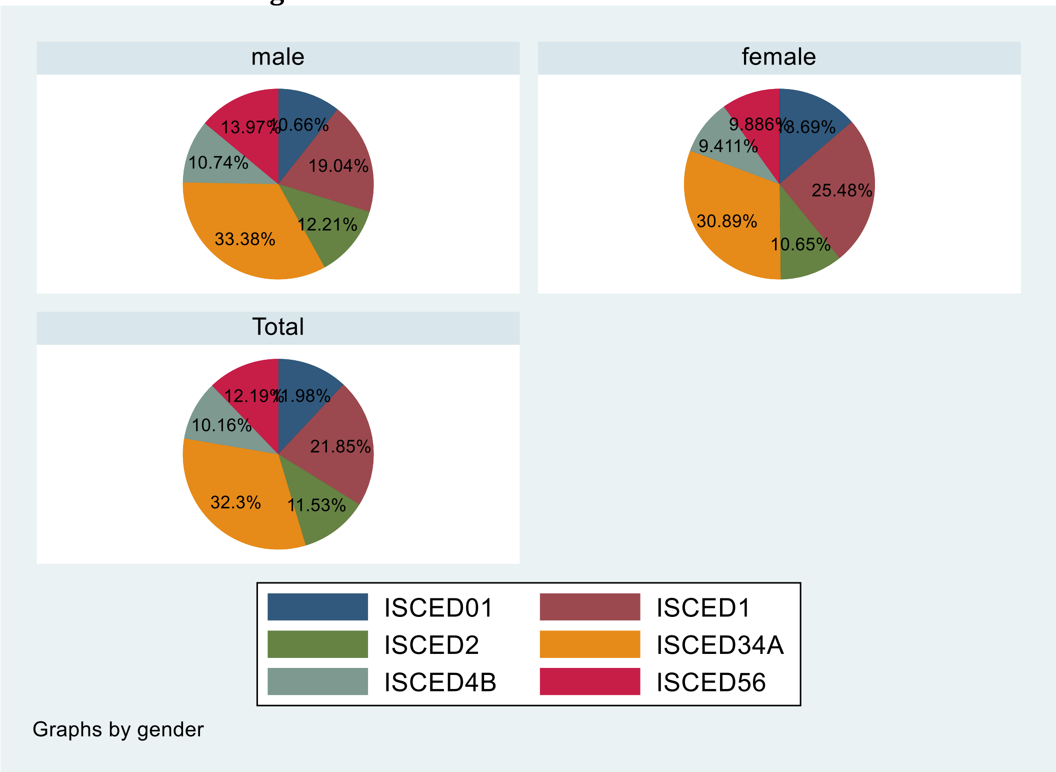
Disaggregating the analytical sample across gender- descriptive evidence based on freq, age and *years_educ_act*

	Freq	% (Freq)	Mean	Mean	Median	SD	Min	Max
			Age	***years_educ_act***				
isced (TOTAL)								
Male	1,360	56.38	32.058	10.842	12	4.477	0	22
Female	1,052	43.62	31.395	10.06	11	4.446	0	20
Total	2,412	100	31.769	10.505	12	4.479	0	22

Source: Author's elaboration of the STEP Household Survey for Kenya. World Bank

Half of the females have under 11 (12 for their male counterparts) years of schooling and the other half have over 11 (12 for their male counterparts) years of schooling, suggestive of the median average for both genders. The maximum years of schooling of 20 (and 22 for males) for females further indicate (beyond the median average) that the males have more years of schooling relative to the females, this suggests the males have higher educational attainment among the employed in urban Kenya.

Disaggregating analytical sample across gender – descriptive evidence based on frequency across credential categories.



Source: Author's elaboration of the STEP Household Survey for Kenya. World Bank

The evidence shows 13.97% (with 9.87% of females) of males make the highest qualified (based on the *isced* credential categories) workforce in Kenya. This indicates that about 190 (with 104 females) males in urban Kenya have credentials from tertiary institutions, representing 86 males over the number of females. Also 145 (with 144 females) males make up those with no formal qualifications, representing a male over the number of females. It is important to note that some of these have some formal years of schooling. Evidence suggests there are (significantly) more qualified males and slightly less qualified males relative to females in urban Kenya, this is subject to further analysis in the analytical chapters.

Informal

Disaggregating analytical sample (all employed) across the *informal* reveals informal workforce makes up 75:13% (with the formal workforce only making 24.87%) of the employed, in urban Kenya; on average, the informal workforce is about five months younger than the formal workforce. Evidence also shows that the mean years of schooling for the formal (13.42) is significantly higher than those of the informal (9.55) with a difference of almost 4 *years_educ_act*. This may give further insights into factors that explain the informality of employment, in urban Kenya.

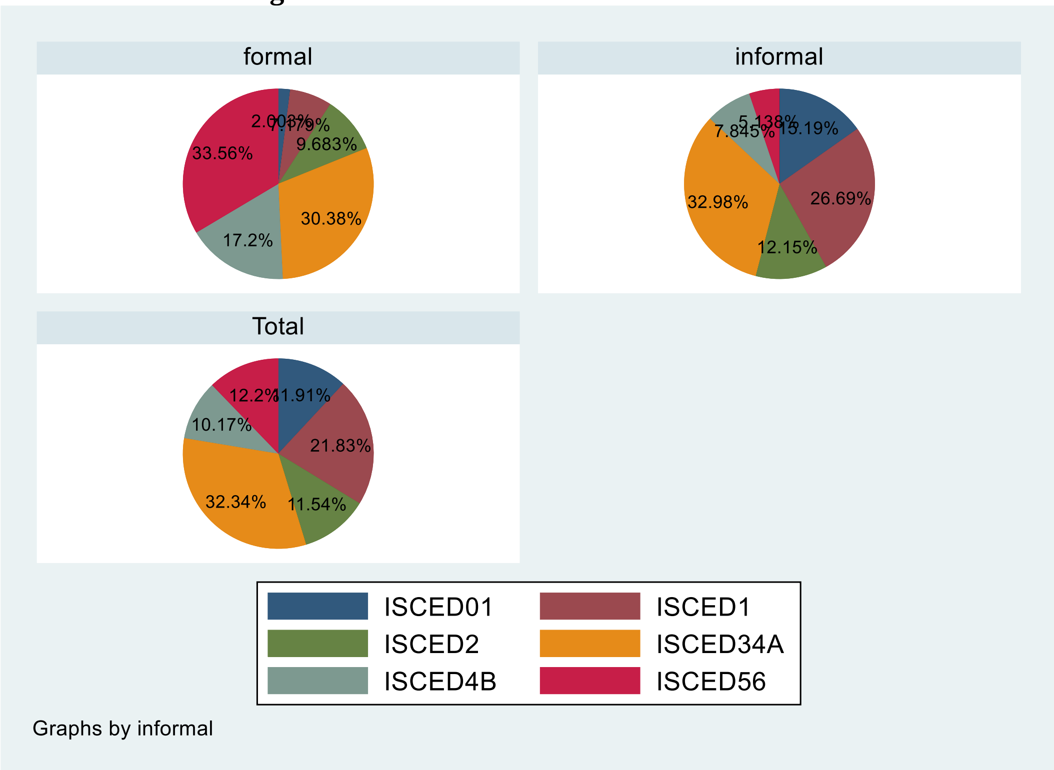
Disaggregating the analytical sample across informal-descriptive evidence based on freq, age and years_educ_act

	Frequency	% (Freq)	Mean	Mean	Median	SD	Min	Max
			age	***years_educ_act***				
Formal	599	24.870	32.097	13.419	14	3.279	0	22
Informal	1,810	75.130	31.652	9.553	10	4.397	0	21
Total	2,409	100	31.763	10.514	12	4.471	0	22

Source: Author's elaboration of the STEP Household Survey for Kenya. World Bank

The significantly high educational attainment of the formal is further accentuated by their median average, with half of the formal having over 14 (with only 10 for the informal workforce) years of schooling and the other half, having under 14 (with only 10 for the informal workforce) years of schooling. The maximum years of schooling of 22 (and 21 for the informal) for the formal further indicate that the formal have higher (significantly) educational attainment relative to the informal workforce in urban Kenya.

Disaggregating analytical sample across informal – descriptive evidence based on frequency across credential categories.



Source: Author's elaboration of the STEP Household Survey for Kenya. World Bank

The evidence shows that 33.56% (with 5.138% of the informal) of the formal making the highest qualified (based on the *isced* credential categories) workforce in Kenya. This indicates that about 188 (with just 93 informal) formal in urban Kenya have credentials from tertiary institutions, representing 95 highly qualified formal workers over the number of informal workers that are highly qualified. Only 12 (with 275 informal) formal workers make up those with no formal qualifications, representing 263 informal workers (with no credentials) over the number of formal workers without credentials. It is important to note that some of these workers (with no formal qualifications) may have some formal years of schooling and may have participated in training, apprenticeships, and certificate programmes, including early childhood education. Overall, evidence from summary statistics suggests there are significantly more (and less) qualified formal (informal) workers; and significantly fewer unqualified (with no credentials) formal workers, relative to the informal workforce in urban Kenya.

Age Groups

Evidence from the Table below shows, disaggregating analytical sample (all employed) across age groups reveals the 25-34 *age_group* make 44.49% of the employed in urban

Kenya; on average, the 25-34 *age_group* is about 28.7 years old on the average, this indicates a young workforce, with an average of 11.2 years of schooling. The 25-34 *age_group*, also has the highest mean years of schooling, relative to other age groups. The evidence suggests, the 15-19 *age_group*, with an average of 8.2 years of schooling, and a mean age of 18.2 years, makes up 2.9% of the workforce in urban Kenya.

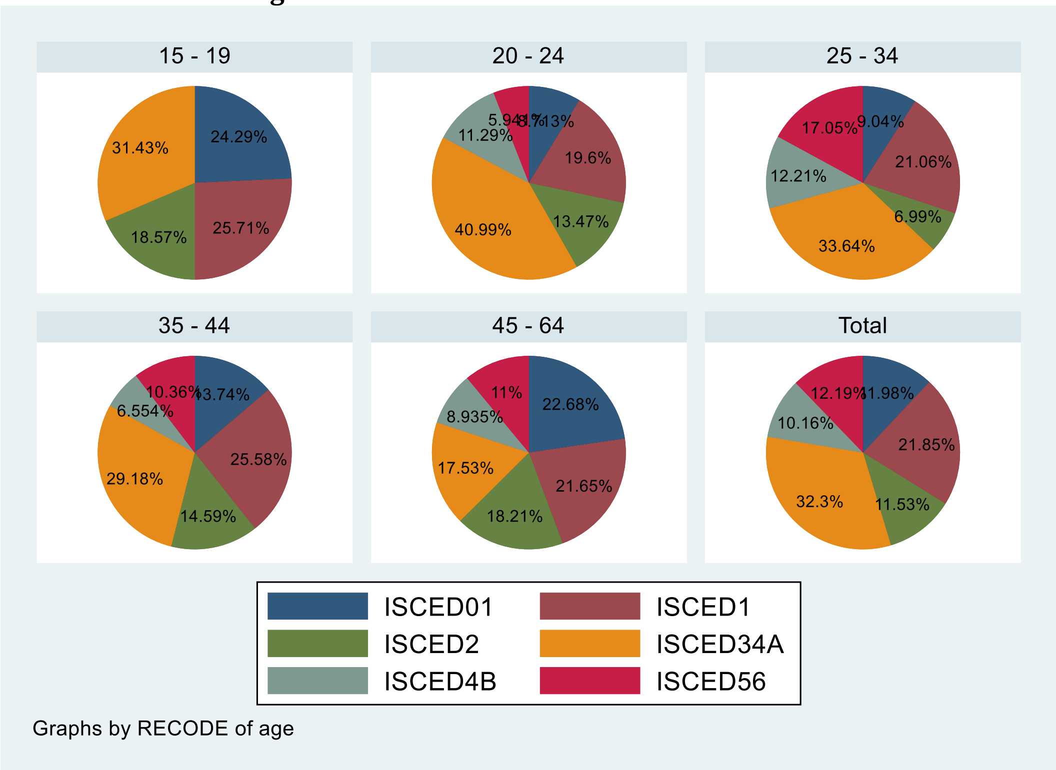
Disaggregating the analytical sample across age_group-
descriptive evidence based on freq, age and years_educ_act

	Freq	**% (Freq)**	**Mean**	**Mean**	**Median**	**SD**	**Min**	**Max**
			age	*years_educ_act*				
15-19	70	2.9	18.243	8.171				
20-24	505	20.94	22.436	10.697	9	4.132	0	13
25-34	1,073	44.49	28.675	11.213	12	3.820	0	18
35-44	473	19.61	38.522	10.034	12	4.241	0	19
45-64	291	12.06	51.650	8.883	10	4.606	0	20
Total	2,412	100	31.769	10.505	9	5.488	0	22

Source: Author's elaboration of the STEP Household Survey for Kenya. World Bank

Half of the 25-34 *age_group* have over 12 (9 for the 15-19 *age_group*) years of schooling and the other half have under 12 (9 for the 15-19 *age_group*) years of schooling, suggestive of the median average for both age groups. The maximum years of schooling of 19 (and 13 for the 15-19 *age_group*) for the 25-34 *age_group*. The group with the highest number of years of schooling, is the 45-64 *age_group*, with 22 years of schooling as the maximum but with a median of 9 years of schooling.

Disaggregating analytical sample across age_group – descriptive evidence based on frequency across credential categories.



Source: Author's elaboration of the STEP Household Survey for Kenya. World Bank

The evidence shows, the 25-34 *age_group* is the most qualified (based on the *isced* credential categories) age group in Kenya, with 17.05% within the ISCED56 category of the *isced.* It is understandable that the 15-19 age group are the least qualified as most in this category are expected to be in full time studies. The 45-64 age group are the next following the 25-34 age group, with 12.19% of this category making up the most qualified (those within ISCED56). As earlier discussed, most of the workers in Kenya fall within the ISCED34A category of the *isced*, with 32.3% of the workforce in urban Kenya having upper secondary and non-tertiary qualifications.

Earnings.

Within Module 4 in the STEP Household Survey are useful labour market outcomes that include earnings.

Descriptive evidence of hourly earnings, log hourly earnings and net profits in US Dollars (USD) are briefly discussed, emphasising distributions across employment status, human capital (credential and skills), age, gender and the informal in urban Kenya. Earnings variables discussed are thus:

Variable	Type	Description
earnings_h_usd	Continuous	USD Hourly labour earnings
In_earnings_h_usd	Continuous	USD Log of hourly labour earnings

Selected earnings variables within STEP HS

Earnings – descriptive evidence, specifications, and discussions.

The hourly earnings (*earnings_h_usd)* is a variable that captures the hourly wage rate on main occupations, across the employed. It is important to note that, it is not typical to be in hourly paid employment in Kenya, as most wage employment in Kenya and non-OECDs are salaried (which may have monthly or weekly payment intervals). Hence, the *earnings_h_usd* has been derived from simple computations involving self-reported salaries – annual, monthly, or weekly; and the variables *hours_d* – the average number of hours of work daily in main occupation; and *hours* – number of hours worked last week, in main occupation. The idea of hourly earnings has long been established in Economics, as this accounts (with ease) for useful levels of demand and supply of labour and efficiency. The idea of hourly earnings also accounts for variations or heterogeneity in human capital by variations in wage rates for different categories of workers. The set hourly wage rate (relative to the minimum hourly wage for the unskilled) can easily determine the type of workers attracted and retained in a job. These ideas are founded and supported by personnel and personnel economics today. However, the work of Lazear (2000) argues that researchers have focussed more on effort (supply and demand of labour) rather than sorting (Lazear, 2000) which relates to the variability of human capital. The idea of hourly earnings in the currency dollars (in lieu of the Kenyan Shilling) is to aid the comparability of research outputs with similar research around the world, as earnings in dollars are widely accepted for empirical studies of this kind. Following the key[148] literature, I adopt the hourly earnings, in USD as the outcome of human capital for the employed in urban Kenya.

[148] See Bick, Blandin and Rogerson (2023).

Distribution of hourly earnings (USD) across the wage-, self- and all-employed.

Source: Author's elaboration of the STEP Household Survey for Kenya. World Bank

Further descriptive evidence of analytical sample across educational attainment, gender, informal, age groups, by earnings.

Educational Attainment, Gender, and Earnings for the Employed in Urban Kenya

Educational Attainment and Earnings in Urban Kenya – Descriptive Evidence

	Freq	**%**	**Mean**	**Mean**	**Median**	**SD**	**Min**	**Max**
		Freq	*(Ys)*	*earnings_h_usd*				
isced								
ISCED01	289	11.980	1.315	1.709	1.140	1.877	0.031	16.286
ISCED1	527	21.850	7.886	2.120	1.070	11.893	0.0428	260.571
ISCED2	278	11.530	10.043	3.111	1.629	8.662	0.023	131.336

ISCED34A	779	32.3	12.426	2.886	1.540	5.392	0.071	80.053
ISCED4B	245	10.16	14.608	5.311	3.170	10.781	0.161	143.764
ISCED56	294	12.19	16.157	9.934	5.990	15.067	0.182	159.737
Total	2,412	100	10.505	3.733	1.629	9.801	0.023	260.571

Source: Author's elaboration of the STEP Household Survey for Kenya. World Bank

Evidence from the above is suggestive of a positive relationship between earnings and educational attainment (based on the mean number of years of schooling (*Ys*) and the ISCED categories). The mean and median earnings for the ISCED34A category of the *isced* is \$2.886038/hr and \$5.391941/hr respectively, this happen to be a point that deviates (Table 18) from the trend between the *isced* and the *earnings_h_usd*. Although variability (SD) in earnings within the ISCED34A category is relatively low, the ISCED34A is the credential category of most of the employed in Kenya, this may be because of high supply (without less demand) of the skills/credential or typical personality of this category of workers in urban Kenya. This will be subjected to further analyses as characteristics of other *isced* categories (particularly, ISCED2 and ISCED4B) may be able to explain the deviation of ISCED34A from the trend. The evidence suggests the highest earning respondent (earning \$260.571/hr) is self-employed, within the ISCED1 category, male gender and in the informal workforce. This appears to be an outlier.

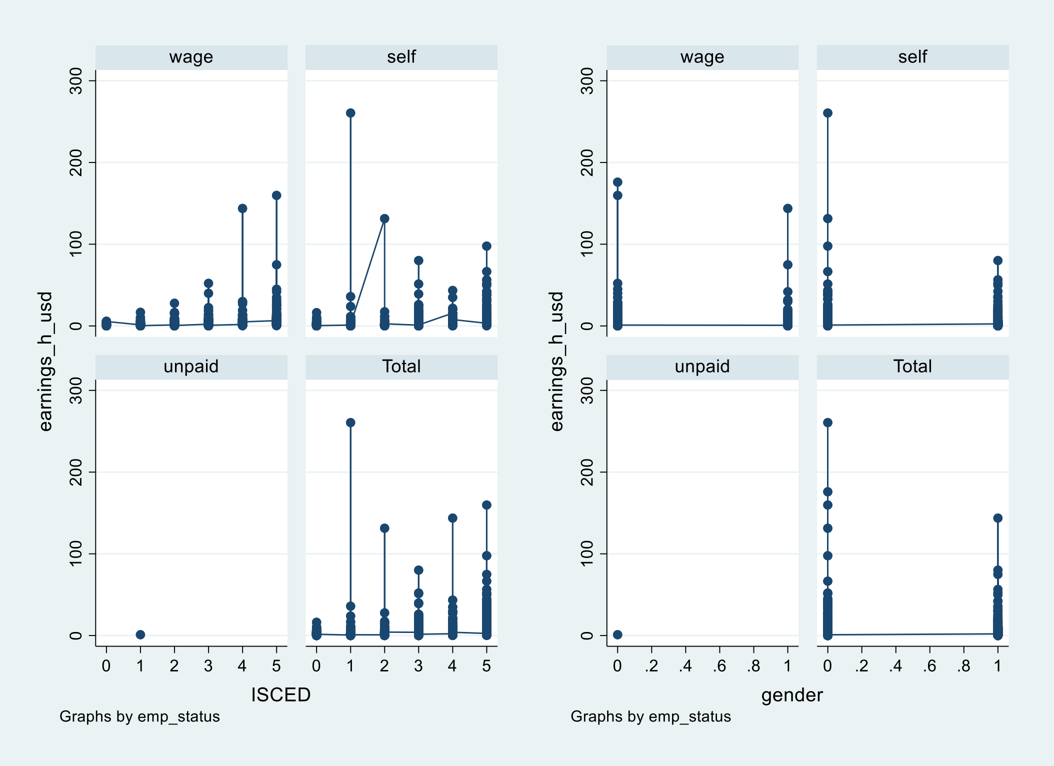
Gender and Earnings in Urban Kenya – Descriptive Evidence

	Freq	**%**	**Mean**	**Mean**	**Median**	**SD**	**Min**	**Max**
		Freq	*(Ys)*	*earnings_h_usd*				
male	1,364	56.32	10.842	4.137	1.797	12.073	0.023	260.571
fem	1,058	43.68	10.068	3.370	1.498	7.730	0.031	143.764
Total	2,422	100	10.505	3.808	1.629	10.437	0.023	260.571

Source: Author's elaboration of the STEP Household Survey for Kenya. World Bank

Evidence from Table 19 and Figure 10 reveals that the males earn $0.77/hr more than females on average, the median earnings for males are also greater, relative to those of the females, although variability (SD) in earnings among the males is greater relative to that of the females. The males earn better than the females across the varying employment categories.

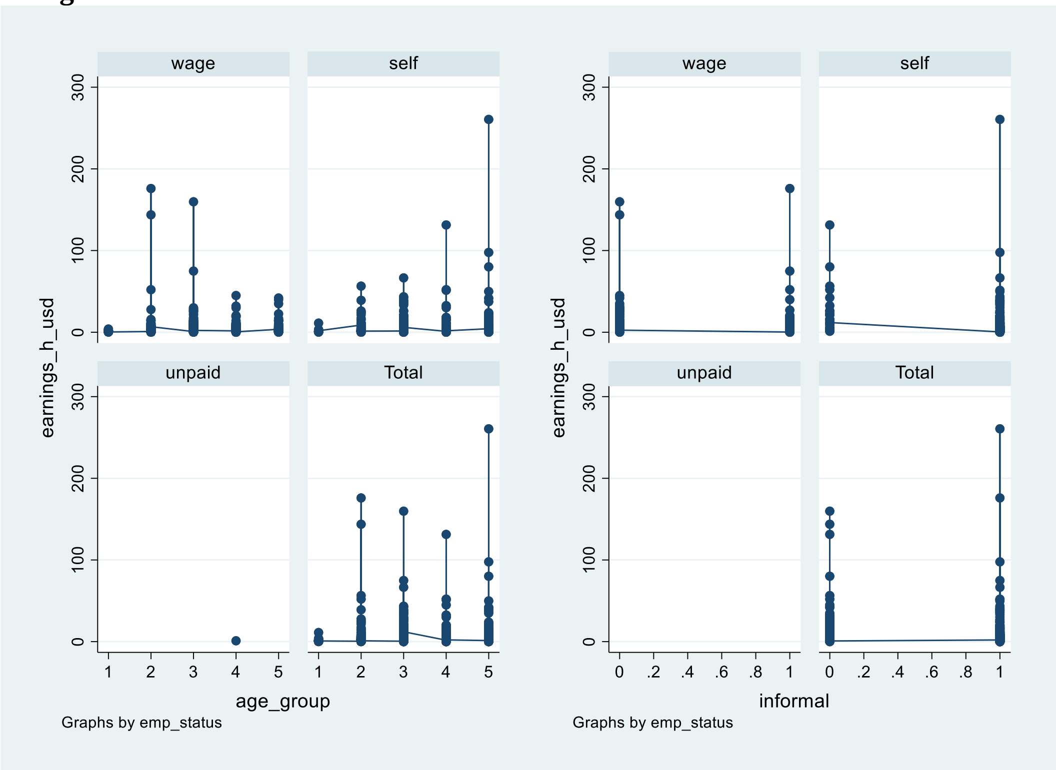
Educational Attainment, Gender, and Earnings – Descriptive Evidence across Employment Category



Source: Author's elaboration of the STEP Household Survey for Kenya. World Bank

ISCED01 – 0 (reference category); ISCED1 – 1; ISCED2 – 2; ISCED34A – 3; ISCED4B – 4; and ISCED56 –5. Gender: 0 – male; 1 – female.

Age Groups, Informal Status and Earnings – Descriptive Evidence across Employment Categories



Source: Author's elaboration of the STEP Household Survey for Kenya. World Bank

age_group:15-19 – 1 (reference category); 20-24 – 2; 25-34 – 3; 35-44 – 4; 45-64 – 5

informal:0 – formal; 1 – informal.

Summary –
Descriptive Evidence of all variables – Analytical Sample 1, with subcategories

Variable	Obs	Mean	Std. dev.	Min	Max
Measures of Human Capital					
Unpaid	67	1.358	1.311	0	4
Self	984	2.182	1.439	0	5
Wage	1,361	2.669	1.551	0	5
isced	***2,412***	***2.434***	***1.529***	***0***	***5***
Unpaid	67	5.851	5.238	0	16
Self	981	8.787	4.779	0	22
Wage	1,356	10.225	4.761	0	22
years_educ	***2,404***	***9.516***	***4.871***	***0***	***22***
Unpaid	67	6.493	5.395	0	15
Self	984	9.881	4.436	0	22
Wage	1,361	11.153	4.293	0	20
years_educ_act	***2,412***	***10.505***	***4.479***	***0***	***22***
Unpaid	67	0.433	0.499	0	1
Self	977	0.692	0.462	0	1
Wage	1,351	0.755	0.430	0	1
ece	***2,395***	***0.720***	***0.449***	***0***	***1***

Unpaid	67	0.134	0.344	0	1
Self	986	0.173	0.379	0	1
Wage	1,369	0.215	0.411	0	1
apprenticeship	***2,422***	***0.196***	***0.397***	***0***	***1***
Unpaid	67	0.0448	0.208	0	1
Self	986	0.069	0.253	0	1
Wage	1,369	0.170	0.375	0	1
training	***2,422***	***0.125***	***0.331***	***0***	***1***
Unpaid	67	0.030	0.172	0	1
Self	986	0.062	0.241	0	1
Wage	1,369	0.103	0.304	0	1
certificate	***2,422***	***0.084***	***0.278***	***0***	***1***
Measures of Cognitive Skills					
Unpaid	64	1.141	0.639	0	3
Self	986	1.263	0.574	0	3
Wage	1,369	1.438	0.809	0	3
num	***2,419***	***1.358***	***0.724***	***0***	***3***
Unpaid	64	0.313	0.664	0	3

Self	986	0.594	1.061	0	3
Wage	1,369	1.151	1.327	0	3
computer	***2,419***	***0.902***	***1.244***	***0***	***3***
Unpaid	64	1.234	1.269	0	3
Self	982	1.518	1.181	0	3
Wage	1,369	1.814	1.172	0	3
read	2,415	1.678	1.189	0	3
Unpaid	64	0.703	0.830	0	3
Self	985	0.912	0.811	0	3
Wage	1,369	1.208	0.989	0	3
write	***2,418***	***1.074***	***0.930***	***0***	***3***
Measures of Personality Traits					
Unpaid	62	2.672	0.437	1.667	3.667
Self	978	2.864	0.580	1.333	4
Wage	1,364	2.857	0.593	1	4
extraversion_av	***2,404***	***2.855***	***0.585***	***1***	***4***
Unpaid	62	3.156	0.486	2	4
Self	978	3.244	0.523	1.667	4

Wage	1,364	3.244	0.498	1.5	4
conscientiousness_av	***2,404***	***3.242***	***0.508***	***1.5***	***4***
Unpaid	62	2.989	0.618	1.667	4
Self	978	2.989	0.576	1	4
Wage	1,364	2.99	0.542	1.333	4
openness_av	***2,404***	***2.992***	***0.558***	***1***	***4***
Unpaid	62	2.710	0.531	1.667	4
Self	978	2.885	0.557	1	4
Wage	1,363	2.841	0.559	1	4
agreeableness_av	***2,403***	***2.855***	***0.558***	***1***	***4***
Unpaid	62	2.710	0.454	1.667	3.667
Self	978	2.685	0.478	1	4
Wage	1,364	2.730	0.506	1	4
stability_av	***2,404***	***2.711***	***0.494***	***1***	***4***
Unpaid	62	2.699	0.559	1.667	4
Self	978	2.774	0.620	1	4
Wage	1,364	2.716	0.584	1	4
grit_av	***2,404***	***2.739***	***0.598***	***1***	***4***

Measures of Earnings					
Unpaid	1	1.070	.	1.070	1.070
Self	904	4.069	12.344	0.031	260.571
Wage	1,330	3.632	8.916	0.023	175.961
earnings_h_usd	***2,235***	***3.808***	***10.437***	***0.023***	***260.571***
Unpaid	1	0.067	.	0.067	0.067
Self	904	0.480	1.202	-3.473	5.563
Wage	1,330	0.652	1.039	-3.771	5.170
ln_earning_h_usd	***2,235***	***0.582***	***1.110***	***-3.771***	***5.563***
Unpaid	1	390.857	.	390.857	390.857
Self	948	637.186	1637.125	0	31268.57
Wage	0				
net_profit_usd	***949***	***636.926***	***1636.281***	***0***	***31268.57***
Employment Categories/Samples					
Unpaid	67	1	0	1	1
Self	986	1	0	1	1
Wage	1,369	1	0	1	1
lm_status	***2,422***	***1***	***0***	***1***	***1***

Unpaid	67	3	0	3	3
Self	986	2	0	2	2
Wage	1,369	1	0	1	1
emp_status	***2,422***	***1.462***	***0.551***	***1***	***3***
Unpaid	67	0	0	0	0
Self	986	0	0	0	0
Wage	1,369	1	0	1	1
wage_worker	***2,422***	***0.565***	***0.496***	***0***	***1***
Unpaid	0				
Self	0				
Wage	1,358	0.119	0.324	0	1
pub_emp	***1,358***	***0.119***	***0.324***	***0***	***1***
Unpaid	67	0	0	0	0
Self	986	1	0	1	1
Wage	1,369	0	0	0	0
self_emp	***2,422***	***0.407***	***0.491***	***0***	***1***
Unpaid	1	1	.	1	1
Self	982	0.829	0.377	0	1

Wage	0				
business_size	***983***	***0.829***	***0.377***	***0***	***1***
Unpaid	64	1	0	1	1
Self	986	0.969	0.175	0	1
Wage	1,369	0.582	0.493	0	1
informal	***2,419***	***0.751***	***0.433***	***0***	***1***
Unpaid	64	3.110	1.416	1	5
Self	986	2.201	0.774	1	5
Wage	1,369	2.271	1.058	0	5
occtype_step	***2,419***	***2.265***	***0.975***	***0***	***5***
Controls: Childhood and Background Characteristics					
Unpaid	67	0.522	0.503	0	1
Self	986	0.523	0.500	0	1
Wage	1,369	0.370	0.483	0	1
gender	***2,422***	***0.437***	***0.496***	***0***	***1***
Unpaid	67	31.821	13.200	18	60
Self	986	32.967	9.944	15	64
Wage	1,369	30.871	9.131	16	64

age	***2,422***	***31.750***	***9.650***	***15***	***64***
Unpaid	65	49.477	48.248	0	300
Self	986	58.183	61.044	0	408
Wage	1,368	51.896	60.760	0	528
tenure	***2,419***	***54.393***	***60.636***	***0***	***528***
Unpaid	63	1.873	0.635	1	3
Self	980	1.847	0.604	1	3
Wage	1,364	1.858	0.563	1	3
ses	***2,407***	***1.854***	***0.582***	***1***	***3***
Unpaid	64	1.156	1.428	0	7
Self	986	1.178	1.490	0	9
Wage	1,369	1.072	1.402	0	9
shocks	***2,419***	***1.117***	***1.440***	***0***	***9***
Unpaid	67	0.060	0.239	0	1
Self	986	0.058	0.233	0	1
Wage	1,369	0.040	0.195	0	1
chronic	***2,422***	***0.048***	***0.213***	***0***	***1***
Unpaid	67	1.284	0.623	0	2

Self	986	1.069	0.468	0	2
Wage	1,369	1.088	0.458	0	2
father_mother	***2,422***	***1.086***	***0.468***	***0***	***2***
Unpaid	67	1.716	1.056	0	4
Self	986	1.420	1.150	0	4
Wage	1,369	1.609	1.175	0	4
max_parent_educ	***2,422***	***1.535***	***1.165***	***0***	***4***
School Characteristics					
Unpaid	44	1.546	0.504	1	2
Self	942	1.656	0.506	1	3
Wage	1,321	1.628	0.517	1	3
school_location	***2,307***	***1.638***	***0.512***	***1***	***3***
Unpaid	44	1.182	0.582	1	3
Self	942	1.331	0.707	1	3
Wage	1,329	1.290	0.653	1	3
school_type	***2,315***	***1.305***	***0.675***	***1***	***3***

Source: Author's elaboration of the STEP Household Survey for Kenya. World Bank

List of References

Aakvik, A., Salvanes, K.G. and Vaage, K., 2010. Measuring heterogeneity in the returns to education using an education reform. *European Economic Review*, *54*(4), pp.483-500.

Abadie, A., Athey, S., Imbens, G.W. and Wooldridge, J.M., 2023. When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, *138*(1), pp.1-35.

Acemoglu, D. and Angrist, J., 2000. How large are human-capital externalities? Evidence from compulsory schooling laws. *NBER macroeconomics annual*, *15*, pp.9-59.

Acemoglu, D. and Autor, D., 2011. Skills, tasks, and technologies: Implications for employment and earnings. In *Handbook of labor economics* (Vol. 4, pp. 1043-1171). Elsevier.

Achoka, J.S., Odebero, S.O., Maiyo, J.K. and Maluko, N.J., 2007. Access to basic education in Kenya. *Inherent Concerns, Education Research and Review*, *2*(10), pp.275-284.

Alderman, H., and King, E.M., 1998. Gender differences in parental investment in education. *Structural Change and Economic Dynamics*, *9*(4), pp.453-468.

Alesina, A., Hohmann, S., Michalopoulos, S. and Papaioannou, E., 2021. Intergenerational mobility in Africa. *Econometrica*, *89*(1), pp.1-35.

Almond, D., Currie, J. and Duque, V., 2018. Childhood circumstances and adult outcomes: Act II. *Journal of Economic Literature*, *56*(4), pp.1360-1446.

Angrist, J.D. and Imbens, G.W., 1995. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association*, *90*(430), pp.431-442.

Angrist, J.D., and Newey, W.K., 1991. Over-identification tests in earnings functions with fixed effects. *Journal of Business & Economic Statistics*, *9*(3), pp.317-323.

Angrist, J.D. and Pischke, J.S., 2009. Instrumental variables in action: sometimes you get what you need. *Mostly harmless econometrics: an empiricist's companion*, pp.113-220.

Angrist, J.D. and Pischke, J.S., 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Angrist, J.D., Imbens, G.W. and Rubin, D.B., 1996. Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, *91*(434), pp.444-455.

ANGRIST, J.D., and KRUEGER, A.B., 1991. QUARTERLY JOURNAL. *The Quarterly-Journal of Economics*.

Anker, R., 1998. *Gender and jobs: Sex segregation of occupations in the world*. International Labour Organization.

Aparicio, J., Cordero, J.M. and Ortiz, L., 2022. Plausible values and their use in efficiency analyses with educational data. *Applied Economics*, *54*(29), pp.3340-3352.

Appleton, S., Bigsten, A. and Manda, D., 1999. Educational expansion and economic decline: returns to education in Kenya, 1978-1995.

Araki, S., 2020. Educational expansion, skills diffusion, and the economic value of credentials and skills. *American Sociological Review*, *85*(1), pp.128-175.

Autor, D.H. and Dorn, D., 2013. The growth of low-skill service jobs and the polarization of the US labor market. *American economic review*, *103*(5), pp.1553-1597.

Autor, D.H., Dorn, D., and Hanson, G.H., 2015. Untangling trade and technology: Evidence from local labour markets. *The Economic Journal*, *125*(584), pp.621-646.

Azfar, O. and Murrell, P., 2009. Identifying reticent respondents: assessing the quality of survey data on corruption and values. *Economic development and cultural change*, *57*(2), pp.387-411.

Azomahou, T.T. and Yitbarek, E.A., 2016. Intergenerational education mobility in Africa: Has progress been inclusive? *World Bank Policy Research Working Paper*, (7843).

Baker, B.D. and Green III, P.C., 2009. Equal educational opportunity and the distribution of state aid to schools: Can or should school racial composition be a factor? *Journal of Education Finance*, pp.289-323.

Baker, B.D. and Green, P.C., 2014. The politics of school finance in the new normal era. In *Handbook of Education Politics and Policy* (pp. 166-188). Routledge

Barone, C., and Van de Werfhorst, H.G., 2011. Education, cognitive skills, and earnings in comparative perspective. *International Sociology*, *26*(4), pp.483-502.

Barro, R.J., and Lee, J.W., 2013. A new data set of educational attainment in the world, 1950–2010. *Journal of development economics*, *104*, pp.184-198.

Barro, R.J., 1991. Economic growth in a cross section of countries. *The quarterly journal of economics*, *106*(2), pp.407-443.

Bassi, M., Busso, M., Urzúa, S. and Vargas, J., 2012. *Disconnected: Skills, education, and employment in Latin America*. Inter-American Development Bank.

Becker, G.S. and Tomes, N., 1986. Human capital and the rise and fall of families. *Journal of labor economics*, *4*(3, Part 2), pp. S1-S39.

Behrman, J.R. and Knowles, J.C., 1999. Household income and child schooling in Vietnam. *The World Bank Economic Review*, *13*(2), pp.211-256.

Benhabib, J. and Spiegel, M.M., 1994. The role of human capital in economic development evidence from aggregate cross-country data. *Journal of Monetary economics*, *34*(2), pp.143-173.

Bellemare, M.F., 2020. How to write Applied Papers in Economics. *Recuperado de http://marcfbellemare. com/wordpress/13712*.

Berman, E., Bound, J. and Machin, S., 1998. Implications of skill-biased technological change: international evidence. *The quarterly journal of economics*, *113*(4), pp.1245-1279.

Bibby, Y., 2020. *Plausible Values: How Many for Plausible Results* (Doctoral dissertation, University of Melbourne).

Björklund, A. and Salvanes, K.G., 2011. Education and family background: Mechanisms and policies. In *Handbook of the Economics of Education* (Vol. 3, pp. 201-247). Elsevier.

Black, S.E. and Devereux, P.J., 2011. Recent developments in intergenerational mobility. *Handbook of labor economics*, *4*, pp.1487-1541.

Black, S.E., Devereux, P.J. and Salvanes, K.G., 2005. The more the merrier? The effect of family size and birth order on children's education. *The Quarterly Journal of Economics*, *120*(2), pp.669-700.

Blanchflower, D. and Elias, P., 1999. Ability, schooling, and earnings: Are twins different? *Unpublished manuscript, Warwick University*.

Bick, A., Blandin, A. and Rogerson, R., 2023. Hours Worked and Lifetime Earnings Inequality

Blinder, A.S., 1973. Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, pp.436-455.

Bound, J., Jaeger, D.A., and Baker, R.M., 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American statistical association*, *90*(430), pp.443-450.

Borghans, L., Green, F. and Mayhew, K., 2001. Skills measurement and economic analysis: an introduction. *Oxford Economic Papers*, *53*(3), pp.375-384.

Branson, N., Garlick, J., Lam, D. and Leibbrandt, M., 2012. Education and Inequality: The South African Case.

Brot-Goldberg, Z.C., Chandra, A., Handel, B.R. and Kolstad, J.T., 2017. What does a deductible do? The impact of cost-sharing on health care prices, quantities, and spending dynamics. *The Quarterly Journal of Economics*, *132*(3), pp.1261-1318.

Brunello, G. and Schlotter, M., 2011. Non-cognitive skills and personality traits: Labour market relevance and their development in education & training systems.

Campbell, F., Conti, G., Heckman, J.J., Moon, S.H., Pinto, R., Pungello, E. and Pan, Y., 2014. Early childhood investments substantially boost adult health. *Science*, *343*(6178), pp.1478-1485.

Card, D., and Krueger, A.B., 1992. Does school quality matter? Returns to education and the characteristics of public schools in the United States. *Journal of political Economy*, *100*(1), pp.1-40.

Card, D., and Krueger, A.B., 1992. School quality and black-white relative earnings: A direct assessment. *The quarterly journal of Economics*, *107*(1), pp.151-200.

Card, D., 1993. Using geographic variation in college proximity to estimate the return to schooling.

Card, D., Cardoso, A.R., Heining, J. and Kline, P., 2018. Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics*, *36*(S1), pp. S13-S70.

Carlsson, M., Dahl, G.B., Öckert, B. and Rooth, D.O., 2015. The effect of schooling on cognitive skills. *Review of Economics and Statistics*, *97*(3), pp.533-547.

Carneiro, P., Crawford, C. and Goodman, A., 2007. The impact of early cognitive and non-cognitive skills on later outcomes.

Carneiro, P., Heckman, J.J. and Vytlacil, E.J., 2011. Estimating marginal returns to education. *American Economic Review*, *101*(6), pp.2754-2781.

Carneiro, P.M. and Heckman, J.J., 2003. Human capital policy.

Case, A. and Deaton, A., 1999. School inputs and educational outcomes in South Africa. *The Quarterly Journal of Economics*, *114*(3), pp.1047-1084.

Cattaneo, M.A. and Wolter, S.C., 2015. Better migrants, better PISA results: Findings from a natural experiment. *IZA Journal of Migration*, *4*(1), p.18.

Checchi, D., Fiorio, C.V. and Leonardi, M., 2013. Intergenerational persistence of educational attainment in Italy. *Economics letters*, *118*(1), pp.229-232.

Chevalier, A., Harmon, C., Walker, I. and Zhu, Y., 2004. Does education raise productivity, or just reflect it? *The Economic Journal*, *114*(499), pp. F499-F517.

Chevalier, A., Harmon, C., O'Sullivan, V. and Walker, I., 2013. The impact of parental income and education on the schooling of their children. *IZA Journal of Labor Economics*, *2*(1), pp.1-22.

Chicoine, L., 2012. Education and fertility: Evidence from a policy change in Kenya.

Cobb-Clark, D.A., and Nguyen, H.T., 2010. Immigration background and the intergenerational correlation in education. *Available at SSRN 1618644*.

Cohen, D. and Soto, M., 2007. Growth and human capital: good data, good results. *Journal of economic growth*, *12*, pp.51-76.

Cohen, D. and Soto, M., 2007. *Growth and education: Good data, good results* (No. halshs-00754251). HAL.

Coleman, J.S., 1966. Equal schools or equal students? *The Public Interest*, *4*, p.70.

Coleman, J.S., 1966. *Equality of educational opportunity [summary report* (Vol. 1). US Department of Health, Education, and Welfare, Office of Education.

Corral, P., Dehnen, N., D'Souza, R., Gatti, R. and Kraay, A., 2021. The world bank human capital index. In *Measuring Human Capital* (pp. 55-81). Academic Press.

Cunha, F. and Heckman, J., 2007. The technology of skill formation. *American economic review*, *97*(2), pp.31-47.

Cunha, F., and Heckman, J.J., 2008. Formulating, identifying, and estimating the technology of cognitive and noncognitive skill formation. *Journal of human resources*, *43*(4), pp.738-782.

Cunha, F., Heckman, J.J. and Schennach, S.M., 2010. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, *78*(3), pp.883-931.

Dee, T.S., 2004. Are there civic returns to education? *Journal of public economics*, *88*(9-10), pp.1697-1720.

Dickson, M. and Smith, S., 2011. What determines the return to education: an extra year or a hurdle cleared? The *Economics of Education Review*, *30*(6), pp.1167-1176.

Downes, T.A., 1992. Evaluating the impact of school finance reform on the provision of public education: The California case. *National Tax Journal*, *45*(4), pp.405-419.

Dustmann, C. and Schoenberg, U., 2008. Why does the German apprenticeship system work? S*kill formation: interdisciplinary and cross-national perspectives*.

Eeckhout, J., Pinheiro, R. and Schmidheiny, K., 2014. Spatial sorting. *Journal of Political Economy*, *122*(3), pp.554-620.

Fitzenberger, B., Mena, G., Nimczik, J. and Sunde, U., 2022. Personality traits across the life cycle: Disentangling age, period, and cohort effects. *The Economic Journal*, *132*(646), pp.2141-2172.

Freeman, R.B., 1975. Overinvestment in college training? *Journal of human resources*, pp.287-311.

Freeman, R.B., 1986. Demand for education. *Handbook of labor economics*, *1*, pp.357-386

Gabay-Egozi, L. and Yaish, M., 2019. Intergenerational educational mobility and life course earnings in Israel. *Social Science Research*, *83*, p.102302

Gaelle, P., Sanchez Puerta, M.L., Valerio, A. and Rajadel, T., 2014. *STEP Skills Measurement Surveys. Innovative Tools for Assessing Skills. Social Protection and Labor* (No. 1421). Discussion Paper.

Ganzach, Y., 2000. Parents' education, cognitive ability, educational expectations, and educational attainment: Interactive effects. *British Journal of Educational Psychology*, *70*(3), pp.419-441.

Glewwe, P. and Kremer, M., 2006. Schools, teachers, and education outcomes in developing countries. *Handbook of the Economics of Education*, *2*, pp.945-1017

Glewwe, P. ed., 2013. *Education policy in developing countries*. University of Chicago Press.

Glewwe, P., Miguel, E., Schultz, T.P. and Strauss, J., 2008. Handbook of Development Economics

Green, A., Green, F. and Pensiero, N., 2015. Cross-country variation in adult skills inequality: Why are skill levels and opportunities so unequal in anglophone countries? *Comparative Education Review*, *59*(4), pp.595-618.

Griliches, Z., 1977. Estimating the returns to schooling: Some econometric problems. *Econometrica: Journal of the Econometric Society*, pp.1-22.

Günther, I. and Launov, A., 2012. Informal employment in developing countries: Opportunity or last resort? *Journal of development economics*, *97*(1), pp.88-98.

Guthrie, J.W. and Wong, K.K., 2012. Education finance from the perspective of politics, political cultures, and government. In *Handbook of research in education finance and policy* (pp. 83-99). Routledge.

Hall, R.E. and Jones, C.I., 1999. Why do some countries produce so much more output per worker than others? *The quarterly journal of economics*, *114*(1), pp.83-116.

Hanushek, E. and Woessmann, L., 2015. The economic impact of educational quality. *Handbook of international development and education*, *2*.

Hanushek, E.A. and Luque, J.A., 2003. Efficiency and equity in schools around the world. *Economics of education Review*, *22*(5), pp.481-502.

Hanushek, E.A. and Woessmann, L., 2008. The role of cognitive skills in economic development. *Journal of economic literature*, *46*(3), pp.607-668.

Hanushek, E.A. and Wößmann, L., 2007. The role of school improvement in economic development.

Hanushek, E.A. and Zhang, L., 2009. Quality-consistent estimates of international schooling and skill gradients. *Journal of Human capital*, *3*(2), pp.107-143.

Hanushek, E.A., 1986. The economics of schooling: Production and efficiency in public schools. *Journal of economic literature*, *24*(3), pp.1141-1177.

Hanushek, E.A., 2003. The failure of input-based schooling policies. *The economic journal*, *113*(485), pp. F64-F98.

Hanushek, E.A., 2013. Economic growth in developing countries: The role of human capital. *Economics of education review*, *37*, pp.204-212.

Hanushek, E.A., Schwerdt, G., Wiederhold, S. and Woessmann, L., 2015. Returns to skills around the world: Evidence from PIAAC. *European Economic Review*, *73*, pp.103-130.

Hanushek, E.A., 2020. Education production functions. In *The economics of education* (pp. 161-170). Academic Press.

Hanushek, E.A., 2010. Education production functions: Evidence from developed countries. *Economics of education*, *2*, pp.132-136.

Hanushek, E., 1970. The production of education, teacher quality, and efficiency. *US Office of Education, Do Teachers Make a Difference*, pp.79-99.

Hanushek, E.A., 1986. The economics of schooling: Production and efficiency in public schools. *Journal of economic literature*, *24*(3), pp.1141-1177.

Harmon, C. and Walker, I., 1995. Estimates of the economic return to schooling for the United Kingdom. *The American Economic Review*, *85*(5), pp.1278-1286.

Harmon, C., Walker, I. and Westergaard-Nielsen, N., 2001. Returns to education in Europe. *Public funding and private returns to education*.

Harmon, Colm, Hessel Oosterbeek, and Ian Walker. "The returns to education: Microeconomics." *Journal of economic surveys* 17, no. 2 (2003): 115-156.

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

Heckman, J.J., Humphries, J.E. and Veramendi, G., 2018. Returns to education: The causal effects of education on earnings, health, and smoking. *Journal of Political Economy*, *126*(S1), pp. S197-S246.

Hedges, L.V., Laine, R.D. and Greenwald, R., 1994. Money does matter somewhere: A reply to Hanushek. *Educational Researcher*, *23*, pp.9-9.

Hertz, T. and Jayasundera, T., 2007. School construction and intergenerational mobility in Indonesia. *American University, Washington DC Department of Economics Working Paper Series*, (2007-18).

Hertz, T., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N. and Veraschagina, A., 2007. Cross-Country Differences and Trends in Intergenerational Socioeconomic Mobility. *Unpublished paper. American University. February*.

Ingram, B.F., and Neumann, G.R., 2006. The returns to skill. *Labour economics*, *13*(1), pp.35-59.

Inyega, J.O., Arshad-Ayaz, A., Naseem, M.A., Mahaya, E.W. and Elsayed, D., 2021, May. Post-independence basic education in Kenya: An historical analysis of curriculum reforms. In *FIRE: Forum for International Research in Education* (Vol. 7, No. 1, pp. 1-23).

Karalashvili, N., Kraay, A. and Murrell, P., 2018. *Doing the survey two-step: The effects of reticence on estimates of corruption in two-stage survey questions* (pp. 335-387). Springer International Publishing.

Karp, M.M., 2015. Dual enrollment, structural reform, and the completion agenda. *New directions for community colleges*, *2015*(169), pp.103-111.

Kautz, T., Heckman, J.J., Diris, R., Ter Weel, B. and Borghans, L., 2014. Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success.

Kimenyi, M.S., Mwabu, G. and Manda, D.K., 2006. Human capital externalities and private returns to education in Kenya. *Eastern Economic Journal*, *32*(3), pp.493-513.

Klasen, S., 2019. What explains uneven female labor force participation levels and trends in developing countries? *The World Bank Research Observer*, *34*(2), pp.161-197.

Kline, P., 2011. Oaxaca-Blinder as a reweighting estimator. *American Economic Review*, *101*(3), pp.532-537.

Klugman, J., Hanmer, L., Twigg, S., Hasan, T., McCleary-Sills, J. and Santamaria, J., 2014. *Voice and agency: Empowering women and girls for shared prosperity*. World Bank Publications.

Kraay, A., 2019. The World Bank human capital index: a guide. *The World Bank Research Observer*, *34*(1), pp.1-33.

Krishnakumar, J. and Nogales, R., 2020. Education, skills, and a good job: A multidimensional econometric analysis. *World Development*, *128*, p.104842.

Krueger, A.O., 2002. *Political economy of policy reform in developing countries* (Vol. 4). MIT press.

Kwenda, P., Ntuli, M. and Gwatidzo, T., 2015. Temporal developments in intergenerational transmission of education: Case for black South Africans. *Research in Social Stratification and Mobility*, *42*, pp.96-113.

Laajaj, R. and Macours, K., 2017. *Measuring skills in developing countries*. The World Bank.

Laborda, L., Elosúa, M.R. and Gómez-Veiga, I., 2019. Ethnicity and intelligence in children exposed to poverty environments: An analysis using the Oaxaca-Blinder decomposition. *Intelligence, 72,* pp.49-58.

Laukaityte, I. and Wiberg, M., 2017. Using plausible values in secondary analysis in large-scale assessments. *Communications in statistics-Theory and Methods*, *46*(22), pp.11341-11357.

Laible, M.C. and Brenzel, H., 2022. Does personality matter? Noncognitive skills and the male migrant wage gap in Germany. *International Migration Review*, *56*(2), pp.376-409.

Lazear, E.P., 2000. Performance pay and productivity. *American Economic Review, 90(5),* pp.1346-1361.

Le, T., Gibson, J. and Oxley, L., 2005. Measures of human capital: A review of the literature.

Lebeau, Y. and Oanda, I.O., 2020. *Higher education expansion and social inequalities in sub-Saharan Africa: Conceptual and empirical perspectives* (No. 2020-10). UNRISD Working Paper.

Lee, J.W. and Lee, H., 2016. Human capital in the long run. *Journal of development economics*, *122*, pp.147-169.

Lee, J.W. and Wie, D., 2017. Returns to education and skills in the labor market: Evidence from Japan and Korea. *Asian Economic Policy Review*, *12*(1), pp.139-160.

Levy, S. and Schady, N., 2013. Latin America's social policy challenge: Education, social insurance, redistribution. *Journal of Economic Perspectives, 27*(2), pp.193-218.

Liu, Z., 2007. The external returns to education: Evidence from Chinese cities. *Journal of Urban Economics*, *61*(3), pp.542-564.

Lucas Jr, R.E., 1988. On the mechanics of economic development. *Journal of monetary economics*, *22*(1), pp.3-42.

Lucas, A.M. and Mbiti, I.M., 2012. Access, sorting, and achievement: The short-run effects of free primary education in Kenya. *American Economic Journal: Applied Economics*, *4*(4), pp.226-253.

Machin, S., 2007. Education expansion and intergenerational mobility in Britain. *Schools and the Equal Opportunity Problem. The MIT Press: Cambridge*, *7*, pp.1-63. Monk, D.H., 1990. Educational costs and small rural schools. *Journal of Education Finance*, *16*(2), pp.213-225.

Machin, S., Marie, O. and Vujić, S., 2011. The crime reducing effect of education. *The Economic Journal*, *121*(552), pp.463-484.

Manda, D.K. and Odhiambo, W., 2003. Urban poverty and labour force participation in Kenya.

Manski, C.F., 1993. Dynamic choice in social settings: Learning from the experiences of others. *Journal of Econometrics*, *58*(1-2), pp.121-136.

Medoff, J.L. and Abraham, K.G., 1980. Experience, performance, and earnings. *The Quarterly Journal of Economics*, *95*(4), pp.703-736.

Meghir, C. and Palme, M., 2005. Educational reform, ability, and family background. *American Economic Review*, *95*(1), pp.414-424

Mincer, J., 1958. Investment in human capital and personal income distribution. *Journal of political economy*, *66*(4), pp.281-302.

Mincer, J., 1974. Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2.

Moretti, E., 2004. Human capital externalities in cities. In *Handbook of regional and urban economics* (Vol. 4, pp. 2243-2291). Elsevier.

Moretti, E., 2006. Private and social returns to education. *Rivista di Politica Economica*, *96*(5/6), p.3.

Morgan, S.L. and Winship, C., 2015. *Counterfactuals and causal inference*. Cambridge University Press.

Mosca, I. and Wright, R., 2018. Is Personality Endogenous? Evidence from Ireland.

Mulligan, C.B. and Sala-i-Martin, X., 1997. A labor income-based measure of the value of human capital: An application to the states of the United States. *Japan and the World Economy*, *9*(2), pp.159-191.

Muthen, B., 1983. Latent variable structural equation modeling with categorical data. *Journal of Econometrics*, *22*(1-2), pp.43-65.

Muthén, B., 1984. A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*(1), pp.115-132.

Neal, D.A. and Johnson, W.R., 1996. The role of premarket factors in black-white wage differences. *Journal of political Economy*, *104*(5), pp.869-895.

Nelson, R.R., and Phelps, E.S., 1966. Investment in humans, technological diffusion, and economic growth. *The American economic review*, *56*(1/2), pp.69-75.

Ngware, M.W., Wamukuru, D.K. and Odebero, S.O., 2006. Total quality management in secondary schools in Kenya: extent of practice. *Quality assurance in Education*, *14*(4), pp.339-362.

Nielsen, H.D., 2006. *From schooling access to learning outcomes, an unfinished agenda: an evaluation of World Bank support to primary education*. World Bank Publications

Nimubona, A.D. and Vencatachellum, D., 2007. Intergenerational education mobility of black and white South Africans. *Journal of Population Economics*, *20*, pp.149-182.

Oanda, I.O. and Jowi, J., 2012. University expansion and the challenges to social development in Kenya: Dilemmas and pitfalls. *Journal of Higher Education in Africa/Revue de l'enseignement supérieur en Afrique*, *10*(1), pp.49-71.

Oaxaca, R., 1973. Male-female wage differentials in urban labor markets. *International economic review*, pp.693-709.

Oledibe, 2023a (forthcoming). Educational Attainment and Skills Proliferation in Urban Kenya

Oledibe, 2023b (forthcoming). Private Returns to Education and Skills in Urban Kenya

Oledibe, 2023c (forthcoming). Human Capital Externalities and Social Returns in Urban Kenya

Oledibe, 2023d (forthcoming). Educational Attainment and Returns to Education and Skills in a Dynamic Framework.

Otieno, W. and Colclough, C., 2009. Financing education in Kenya: Expenditures, outcomes, and the role of international aid. *Cheptile Final Document. Retrieved from UON repository: erepository. uonbi. ac. ke*.

Patrinos, H.A., Psacharopoulos, G. and Tansel, A., 2021. Private and social returns to investment in education: the case of Turkey with alternative methods. *Applied Economics*, *53*(14), pp.1638-1658.

Pekkarinen, T., Uusitalo, R. and Kerr, S., 2009. School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform. *Journal of public Economics*, *93*(7-8), pp.965-973.

Pritchett, L., 1997. Divergence, big time. *Journal of Economic perspectives*, *11*(3), pp.3-17.

Pritchett, L., 2001. Where has all the education gone? *The world bank economic review*, *15*(3), pp.367-391.

Psacharopoulos, G. and Arriagada, A.M., 1986. The educational composition of the labour force. *International Labour Review*, *125*(5), pp.561-74

Psacharopoulos, G., 1978. Educational planning: Past and present. *Prospects*, *8*(2), pp.135-142.

Psacharopoulos, G., 1989. Time trends of the returns to education: Cross-national evidence. *Economics of Education Review*, *8*(3), pp.225-231.

Ravallion, M., 2014. Income inequality in the developing world. *Science*, *344*(6186), pp.851-855.

Rahimi, E. and Hashemi Nazari, S.S., 2021. A detailed explanation and graphical representation of the Blinder-Oaxaca decomposition method with its application in health inequalities. *Emerging Themes in Epidemiology*, *18*(1), p.12.

Rice, J.K., Monk, D. and Zhang, J., 2020. School finance: An overview. *The economics of education*, pp.333-344.

Roseth, V.V., Valerio, A. and Gutiérrez, M., 2016. *Education, skills, and labor market outcomes: Results from Large-Scale Adult Skills Surveys in urban areas in 12 countries*. World Bank.

Sacerdote, B., 2011. Peer effects in education: How might they work, how big are they and how much do we know thus far? In *Handbook of the Economics of Education* (Vol. 3, pp. 249-277). Elsevier.

Sánchez, A., 2017. The structural relationship between early nutrition, cognitive skills and non-cognitive skills in four developing countries. *Economics & Human Biology*, *27*, pp.33-54.

Sayed, Y. and Ahmed, R., 2015. Education quality, and teaching and learning in the post-2015 education agenda. *International Journal of Educational Development*, *40*, pp.330-338.

Schiefelbein, E. and Farrell, J.P., 1984. Education and occupational attainment in Chile: The effects of educational quality, attainment, and achievement. *American Journal of Education*, *92*(2), pp.125-162.

Schoellman, T., 2012. Education quality and development accounting. *The Review of Economic Studies*, *79*(1), pp.388-417.

Schultz, T.W., 1961. Investment in human capital. *The American economic review*, *51*(1), pp.1-17.

Serneels, P., 2008. Human capital revisited: The role of experience and education when controlling for performance and cognitive skills. *Labour Economics*, *15*(6), pp.1143-1161.

Shafiq, M.N. and Valerio, A., 2021. The roles of childhood circumstances and schooling on adult reading skills in low-and middle-income countries. *Prospects*, pp.1-20.

Sianesi, B. and Reenen, J.V., 2003. The returns to education: Macroeconomics. *Journal of economic surveys*, *17*(2), pp.157-200.

Skrondal, A. and Rabe-Hesketh, S., 2004. *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Crc Press.

Solon, G., 1999. Intergenerational mobility in the labor market. In *Handbook of labor economics* (Vol. 3, pp. 1761-1800). Elsevier.

Somerset, A., 2007. *A Preliminary Note on Kenya Primary School Enrolment Trends over Four Decades. Create Pathways to Access. Research Monograph No. 9*.

Somerset, A., 2009. Universalising primary education in Kenya: The elusive goal. *Comparative Education, 45*(2), pp.233-250.

Stiglitz, J.E., 2012. *The price of inequality: How today's divided society endangers our future*. WW Norton & Company.

Temple, J., 1999. The new growth evidence. *Journal of economic Literature*, *37*(1), pp.112-156.

Thias, H.H. and Carnoy, M., 1972. Cost-benefit analysis in education: a case study of Kenya. *(No Title)*.

Todd, P.E., and Wolpin, K.I., 2003. On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, *113*(485), pp. F3-F33.

Todd, P.E., and Wolpin, K.I., 2007. The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human capital*, *1*(1), pp.91-136.

Trostel, P., Walker, I. and Woolley, P., 2002. Estimates of the economic return to schooling for 28 countries. *Labour economics, 9*(1), pp.1-16.

Valerio, A., Sanchez Puerta, M.L., Tognatta, N. and Monroy-Taborda, S., 2016. Are there skills payoffs in low-and middle-income countries? empirical evidence using STEP data. *Empirical Evidence Using Step Data (November 3, 2016). World Bank Policy Research Working Paper*, (7879).

Villa, K.M., 2017. Multidimensional human capital formation in a developing country: Health, cognition, and locus of control in the Philippines. *Economics & Human Biology*, *27*, pp.184-197.

Vlasblom, J.D. and Schippers, J.J., 2005. *Arbeidsmarktconsequenties van kinderen krijgen: een verkenning van de langetermijngevolgen*. OSA, Organisatie voor Strategisch Arbeidsmarktonderzoek.

Wambugu, A., 2003. *Essays on earnings and human capital in Kenya* (No. 124)

Widanarko, B., Legg, S., Devereux, J. and Stevenson, M., 2015. Interaction between physical and psychosocial work risk factors for low back symptoms and its consequences amongst Indonesian coal mining workers. *Applied ergonomics*, *46*, pp.158-167.

Widanarko, B., Legg, S., Devereux, J. and Stevenson, M., 2015. Interaction between physical and psychosocial risk factors on the presence of neck/shoulder symptoms and its consequences. *Ergonomics*, *58*(9), pp.1507-1518.

Winship, C. and Morgan, S.L., 1999. The estimation of causal effects from observational data. *Annual review of sociology*, *25*(1), pp.659-706.

Woessmann, L. and Peterson, P.E. eds., 2007. *Schools and the equal opportunity problem*. MIT Press.

Wooldridge, J.M., 2002. Introductory Econometrics: A Modern Approach, 2002. New York: South-Western College Publishing.

Wooldridge, J.M., 2010. *Econometric analysis of cross section and panel data*. MIT press.

Wooldridge, J.M., 2015. Introductory econometrics: A modern approach: Nelson Education. *Scarborough, ON, Canada*.

World Bank, 2018. *World development report 2019: The changing nature of work*. The World Bank.

World Bank, 2018b. The Human Capital Project.