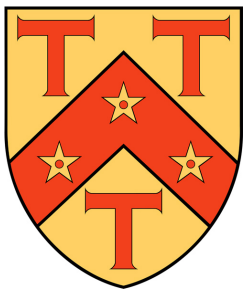


Why Cognitive Skills Matter for Development and What Explains Their Variation:

Evidence from Sudanese National Learning Assessment



Huseyin Atakan Keskin

Advisor: Dr.David Johnson

Department of Education
University of Oxford

A dissertation submitted in partial completion of the
Masters of Science in Comparative and International Education

St. Antony's College August 2019

Note that some graphs/tables/images may be removed in order to comply with copyright restrictions.

I would like to dedicate this thesis to my beloved mother Emine Keskin, who was a kindergarten teacher, a firm believer in quality education, track & field star, and the most loving parent. I hope you are proud ...

Rest in peace ...

Bu tezi, kaliteli eğitime inanan anaokulu öğretmeni, atletizm yıldızı, ve sevgisini asla esirgemeyen değerli annem Emine Keskin'e adıyorum. Umarım gurur duyuyorsundur.

Huzur içinde uyu...

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing, which is the outcome of work done in collaboration with others. This dissertation contains **19,969 words**, excluding cover page, declaration, abstract, appendices, bibliography, table of contents, list of tables, list of figures and nomenclature; and has 18 figures.

Huseyin Atakan Keskin
August 2019

Acknowledgements

English

I would first like to thank **Mr. Kemal Güleryüz**, Chairman of Gülermak Inc, for believing in me and financing a significant portion of my tuition at Oxford. I could not be thankful enough for his support, and I hope to follow his footsteps in the future by helping students to achieve their dreams.

In conjunction with this, I would also like to thank **Ms. Feriha Mert**, who recommended me to Mr. Güleryuz, and is one of the most inspiring women I have ever met who resembles honesty, integrity, hard work, and helping those in need. I would like to thank her for the love and support as she loves me like a son.

This thesis would not have been completed the way it is had I not have the support of **Dr. David F. Johnson**. Sir, thank you so much for encouraging me to take on this thesis. I have grown immensely under your supervision, truly enjoyed working with the best supervisor and mentor. You have treated me not merely as a student, but an economist of education, which helped me to be confident in what I was doing and explore concepts such as machine learning. I look forward to learning from your invaluable knowledge, experiences, and expertise; and contribute together to humanity and science. I hope I can follow your footsteps in the future and bring a profound impact on the lives of those in need like yourself.

I would also like to acknowledge our course leader, **Dr. Maia Chankseliani**. Your work-ethic and expectations from us nudged me to strive to be the best student I can possibly be during this program. You are the epitome of how can hard-work and dedication bring one to the top of the league. I am blessed to be your student and look forward to watching you walk towards a Nobel Prize. With this, I would like to also acknowledge all the other professors and staff in the Department of Education for helping me and creating the best MSc program in the world.

On the other hand, I would like to acknowledge the former President of Colombia/Nobel Peace Laureate **Hon. Mr. Juan Manuel Santos**. Mr. President, thank

you so much for mentoring me through the course of my masters at Oxford. I have learned great lessons from you, and your guidance has helped me to dream bigger and work harder. Thank you for believing in me and showing unwavering support. I hope to follow your footsteps in the future to bring peace and prosperity to the world.

I would like to acknowledge my role model, No. 1 mentor, best friend, my beloved father, **Dr. Kaan Keskin**. You have worked day-and-night to provide me with the best opportunities and setting the groundwork for me to receive the best education in the world. Thank you for allowing me to explore what I truly am the most passionate about and giving me the confidence to make my own decisions. I hope I can be as good of a father as you have been to me to my kids. I am truly blessed to be your son.

I would like to acknowledge **Prof. Dr. Ülkü Sultan Keskin**, the Queen Mother of our house. Along with how loving, supporting, and a great parent she is, she also devotes her life towards contributing to science and providing high-quality education to her students. I would like to thank her for showing me unwavering love and support; also teaching me highly valuable virtues. I hope I, as your son and **Ms. Elif Mira Keskin**, the sunshine of my life, as your daughter will always make you proud and continue to resemble the virtues and great lessons you have been teaching us.

I would like to thank my uncle, big brother, and mentor **Dr. Levent Yılmaz** as I always learn from his valuable experience and knowledge.

I would like to acknowledge my beloved grandmother, **Mrs. Gülşen Yılmaz** for her endless love and support. May she rest in peace.

I would like to thank my loving and supporting grandfather, **Mr. Mustafa Yılmaz**, aka "the Leader" for always being there for me.

I am blessed with a myriad of relatives that have played a critical role in my life. I would like to thank my big family – the members of the **Keskin and Aydoğdu** families. Furthermore, I would like to acknowledge my big auntie **Mrs. Birsen Eroğlu**; my big uncle, **Mr. Oğuz Ünsal Esq.**

I would like to send my most profound respects and regards to **Hon. Mr. Mahmut Nedim Bilgiç** for his service and the vital impact he has brought into the Republic of Turkey; and also for his mentorship and guidance. In this manner, I would like to also thank **Mr. Rafael Hajiyev** for his strategic teachings as I am grateful to learn from his vast knowledge and experience.

I would like to acknowledge my closest friends at Oxford: **Mr. Cole Scanlon**, **Mr. Stefan Baconschi**, **Mr. Jorge Acevedo Piña**, **Mr. Yusuf Oldaç İkbal**, **Mr. Victor Mamoine** ; **Mr. Lavi Lehman**, **Mr. Nijat Hasanli**, **Mr. Abdulla Omaigan**, **Ms. Joanna Li How Cheong**, **Ms. Sıla Karakoç**, **Ms. Jessica**

Anaina, **Ms. Silvia Nuñez**, and all my other friends at the CIE cohort and St. Antony's College.

I would like to thank all my teammates and coaches at **Oxford University Lancers American Football Club**. It was an honor to be a part of this brave team as a linebacker. I will never forget the Varsity Match where we beat Cambridge on their turf after 25 years, 19-9; holding them with zero points for three full quarters. Go, Lancers! Go Blues!

I would like to acknowledge my closest friends all around the world, **Ms. Giovanna Montagnoli**, **Mr. Mert Tepebağ**, **Mr. Furkan Şanlı**, **Mr. Trent Lott**, **Mr. Dom Desire**, **Mr. Bay Downing**, **Mr. Monty Berrow**, **Mr. Oğulcan Uslu**, **Mr. Emre Karahan**, **Mr. Engin Muradoğlu**, **Mr. Can Karaciğan**, **Ms. Güzin Oflaz**, **Mr. Selami Oflaz** , **Ms. Merve Demircan**, **Mr. Musab Ergüneş**, **Mr. Bulut Ünsal**, **Mr. Timur Okyay Oktay**, and **Mr. Berkin Kolot**.

I would like to acknowledge **Mrs. Tuğba Bozçağa**, who is currently reading for a DPhil in Political Science at MIT; along with **Mrs. Gizem Halıs Kasap LLM. Esq.** who is reading for a JSD at Wake Forest University for their unwavering support, teachings, and strategic recommendations for the future trajectory of my career.

I would like to thank all the members of the ***Fair Opportunity Project – Team Turkey***: **Kayra, Beyza, Esra, Melis, and Tuğçe** for all their hard work to provide access to quality education in the world's top universities for Turkish students.

Finally, I would like to thank the College Registrar **Mrs. Filiz McNamara** along with the staff in St. Antony's College; and the Higher Degrees Administrator, **Ms. Kyra Smith**, along with the staff at the Department of Education for their professionalism, help and support.

Teşekkürler.
H.A.K.

Abstract

Growth theories and the role of human capital formation in development remain widely-discussed topics in the field of education economics. While a plethora of studies that use school enrollment rates as a proxy for human capital fail to explain the variation in per-capita GDP growth, more recent work in the economics of education shows that cognitive skills are essential for development. How these skills are formed, however, is less clear and should be investigated. This dissertation examines whether cognitive skills, measured by science and mathematics, explain human development in Sudan. Drawing on the 2018 Sudanese National Learning Assessment, part of the Basic Education Recovery Project of World Bank (2019a), the questions are: Do cognitive skills matter for development? If so, what explains their variation? The findings, achieved through correlation analyses between state-level poverty rates and science & mathematics achievement, show a robust significant relationship. Additionally, multivariate OLS models indicate that components of Cunha and Heckman (2007) theory on skill formation such as subsequent achievement, parental and environmental factors and a student asset index created with a machine-learning algorithm explain the variation in science achievement.

Table of contents

List of figures **xv**

List of tables **xvii**

Nomenclature **xix**

Introduction **1**

1 Growth of Nations **5**

1.1 Eminence of Economic Growth Theories 5

1.1.1 Solow—Swan Model 8

1.2 Human Capital Theory: A Global Phenomenon 10

1.2.1 Education, Human Capital and Economic Growth 11

1.3 Rationalizing the Impact of Education on Economic Growth 13

1.3.1 The Role of Cognitive Skills in Educational Quality 15

1.3.2 Solving the World's Growth Paradox Through Cognitive Skills . 17

1.4 The Impact of SES on Cognitive Skills 23

1.5 Criticism of Human Capital & Measurement Gaps 25

2 Political Economy of Education in Sudan **29**

2.1 Context of the Country 29

2.2 The Economic Performance of Sudan 31

2.2.1 Sudan's Performance in Major Global Indexes 33

2.2.2 Sudan's Growth Accounting 36

2.3 The Status of the Sudanese Education System 37

2.3.1 Historical Context 37

2.3.2 Gross School Enrollment Trends 38

2.3.3 The Governance of Sudanese Education System 40

2.3.4 Key Challenges of the Education System in Sudan 42

3 Methodology **47**
3.1 The Data Set . . . 47
3.1.1 2018 NLA Sample . . . 48
3.1.2 Research Questions . . . 50
3.2 Methods and Analyses . . . 50
3.2.1 Method 1 . . . 50
3.2.2 Method 2 . . . 52
3.2.3 Rigor . . . 55
3.3 Research Ethics . . . 57

4 Results **59**
4.1 Why Cognitive Skills Matter for Development? . . . 59
4.2 What Explains the Variation in Cognitive Skills? . . . 63

5 Discussion and Conclusion **69**
5.1 Further Investigation of Mother's Work Type . . . 69
5.2 Previous Findings vs. the 2018 NLA . . . 70
5.3 Conclusion and Recommendations . . . 72
5.4 Limitations . . . 74

References **77**

Appendix A Further Descriptions & Analyses **87**

Appendix B Codes **101**

List of figures

1.1 An Example of the Global Production Function 7
1.2 Technological Progress and Economic Growth in the Solow Model . . . 9
1.3 Educational Attainment and GDPPC in Venezuela, 1960 – 2000 16
1.4 Knowledge Capital and Economic Growth in the World 18

2.1 Sub-regions, Human Development and Population Growth in Africa . . 30
2.2 Sudan's GDPPC Growth : 1999-2018 32
2.3 2019 Fragile States Index - Africa 34
2.4 2019 Fragile States Index - World 35
2.5 The States of Sudan 38
2.6 School Enrollment Rates by Country and State Level in Sudan 39
2.7 Organization of the Educational Governance in Sudan 40
2.8 Sudanese Public Education Spending by Institutional Level : 2000–2009 41
2.9 Public Education Spending in Africa: 2000–09 43
2.10 Children out of Primary School in Sudan, by State (2016) 44

3.1 2018 Sudanese NLA's Sample Distribution 49

4.1 Correlations between Achievement and Poverty in Sudan 61

5.1 Science Achievement and Mother's Work Type 70
5.2 2009 World Bank Learning Assessment in Math&Reading 71

List of tables

1.1 Hanushek and Woessmann (2015)'s OLS Models 20
1.2 Economic Growth Accounting for Mexico and the US between 1950-2014 using the Human Capital Augmented Solow Model 22

2.1 Economic Growth Accounting for Sudan between 1974-2017 using the Human Capital Augmented Solow Model 36

4.1 Achievement and Educational Spending in Sudan, by State 60
4.2 OLS Models 1-5 64

Nomenclature

Roman Symbols

A	technological progress/solow residual
F	complex function
H	human capital(in function)
h	parental and environmental skills
I	investment in skills
K	capital
L	labor
n	population growth
s	investment
s_h	investment in human capital
s_k	investment in capital
w	weight from the first principal component
Y	output(GDP)
y	GDP per capita (in function)
z	z-value

Greek Symbols

1–α	income share of labor of the total income

$1–\beta$ income share of human capital of the total income

α income share of capital of the total income

β standardized regression coefficient

δ Depreciation

ϵ margin of error

ι unit imaginary number $\sqrt{-1}$

μ mean

π $\simeq 3.14\ldots$

θ stock of skills

Superscripts

$\dot{X}$ derivative of variable X with respect to time

Subscripts

j country j = 1...N

k per effective capital

l per effective labor

t time period

Other Symbols

∂ partial derivative

Acronyms / Abbreviations

AIC Akaike's Info Criterion

C-EUR Central Europe

CDPF Cobb–Douglass Production Function

CLT Central Limit Theorem

COMM Commonwealth OECD members

CPA Comprehensive Peace Agreement

CUREC Central University Research Ethics Committee

DEFF the design effect

DV Dependent Variable

E-ASIA East Asia

EFI Economic Freedom Index

EGMA Early Grade Mathematics Assessment

EGRA Early Grade Reading Assessment

EU The European Union

FMoGE Federal Ministry of General Education

FSI Fragile States Index (2019)

GDP Gross Domestic Product

GDPPC GDP Per Capita

GPE Global Partnership in Education

HC Human Capital

IEA International Association for the Evaluation of Educational Achievement

LATAM Latin America

MENA Middle East and North Africa

MP_K Marginal Product of Capital

MP_L Marginal Product of Labor

N-EUR Northern Europe

NLA National Learning Assessment

OECD Organization for Economic Cooperation and Development

OLS ordinary least squares

PCA Principal Component Analysis

PC Principal Component

PISA Programme for International Student Assessment

PPP Purchasing Power Parity

PWT Penn World Tables

S-EUR Southern Europe

SAI Student Assets Index

SD Standard Deviation

SES socioeconomic status

SSAFR sub-Saharan Africa

TFP Total Factor Productivity

TMC Transitional Military Council

UNESCO The United Nations Educational, Scientific and Cultural Organization

UNICEF United Nations International Children's Emergency Fund

UN United Nations

US United States

VIF Varience Inflation Factors

Introduction

What causes growth and development? Why are some countries more developed than others? These questions have been asked since the beginning of the 20$^{\text{th}}$ century (DeLong, 2000), but yet the answer remains unclear.

We know that there is a myriad of factors that contribute to development: institutional quality and human rights (Acemoglu and Robinson, 2012); economic factors such as capital formation, labor force participation (Cobb and Douglas, 1928) and technological progress (Solow, 1956; Swan, 1956); and other factors which enhance the quality of life. Out of all these factors; however, one might have stood out the most in terms of determining the future trajectory of growth theories and perhaps starting an ever-lasting debate on its potential impact: human capital.

Ever since the first conceptualization of human capital by Smith (1786), and then the advent of the field '*Economics of Education* by a Nobel Laureate Economists such as Schultz (1961) Becker (1962), theories of economic growth via human capital, particularly by using educational variables has been a contentious debate.

Just as backbones of scientific discoveries such as the Higgs boson was confirmed 50 years after its discovery, shedding light on how human capital impacts economic growth also took significant efforts (Acs et al., 2014). The inclusion of human capital into growth models was perspicuous other stocks of economic drivers (Romer, 1990) such as capital and labor were only able to explain a small portion of economic growth (Felipe and Adams, 2005). Measurements of human capital with educational attainment, on the other hand, has brought along significant problems (Barro and Sala-i Martin, 2004) and according to studies of Hanushek and Woessmann (2015) have been an improper measure.

It was until the beginning of the 21st century when Hanushek and Kimko (2000) put forth that the inability of studies to capture the impact of human capital with educational variables on economic growth was due to heavy reliance on the quantity, rather than the quality, of education. Fifteen years later, Hanushek and Woessmann (2015) argued for the usage of educational quality, which is the level of cognitive

skills formed within the educational system of a country, to measure human capital. The authors state that educational quality can be measured by international science and mathematics assessment. Their models explain a significant portion of economic growth rates of 50 countries using science and mathematics achievement as a proxy for cognitive skills.

The critical takeaway of Hanushek and Woessmann (2015)'s study is that "in the long run, a nation's prosperity is directly related to the skills of its population" [p.203], and these skills can be measured by achievement in science and mathematics. The authors show that why despite having high enrollment rates, regions such as Latin America fell behind of regions such as Southeast Asia over the years due to lack of cognitive skill formation. Such finding is also consistent in the works of others such as Acemoglu and Robinson (2012) analyzing development through the quality of institutions. The ultimate notion from the review of the economics of education is that cognitive skills are essential for development, and how these skills are formed should be investigated.

Given the evidence elsewhere, this thesis looks at the relationship between skills and economic development in Sudan. Two interwoven questions are asked: Drawing on the 2018 Sudanese National Learning Assessment (NLA), which is a part of the *Basic Education Recovery Project* of World Bank (2019a), the questions are: Why do cognitive skills matter for development? If so, what explains their variation?

To answer these questions holistically, Chapter **1** starts with an in-depth review of the state of the field in exploring the importance of cognitive skills for development; why cognitive skills should be investigated and the methodological progression of the field. It critically examines how studies have built on previous work that started with Smith (1786)'s conceptualization, Cobb and Douglas (1928), Solow (1956) & Swan (1956)'s theorization and Hanushek and Woessmann (2015) proofs on why human capital, measured with cognitive skills matters for development. Along with discussing criticism and limitations of the human capital approach, it also surveys the studies inquiring how cognitive skills vary.

Chapter **2** analyzes the political economy of education in Sudan as it discusses the context of the country; the performance and historical trends of the economy; economic and social freedoms and the state of the education system. This chapter is specifically essential to form a contextual perspective to interpret the results from the data analyses better.

Chapter **3** starts with a description of the methods used to test the prevalence of claims from Hanushek and Woessmann (2015) on the importance of cognitive skills. It

also takes the field to another methodological progression to analyze the factors that impact cognitive skills, and test one of the most renowned theories on cognitive skills formation by Cunha and Heckman (2007). Alternatively, this chapter also mentions how the analyses in Chapter **4** rely on the critical methods employed in the field, such as machine learning algorithms like the Principle Component Analysis (Taylor and Yu, 2009).

The analyses show essential results. First, the findings, achieved through correlation analyses between state-level poverty rates and science & mathematics achievement, show a robust significant relationship in Section **4.1**. Additional descriptions and analyses, such as the relationship between cognitive skills and types of governmental spending on education, are shown in Appendix **A**.

In Section **4.2**, multivariate OLS models indicate that components of Cunha and Heckman (2007) theory on skill formation such as subsequent achievement, parental and environmental factors and a student asset index created with a machine-learning algorithm explain the variation in science achievement. I finally conclude in Chapter **5** with discussing how the results compare with the literature in Chapter **1**; and limitations of the NLA data along with developing recommendations and research questions for future research.

While some minor analyses from Chapter 1-3 rely on secondary data from leading sources such as the World Bank Indicators, Penn World Tables and others; the primary data used in Chapter **4** that this thesis is built upon comes as a secondary source from Johnson (2018).

I use *R* Programming Software (R Core Team, 2019) for data analysis and plot creation; and use Phyton Programming Software for data science purposes such as data merging and performing machine learning algorithms, which the codes written for such purposes are shown in Appendix **B**.

Chapter 1

Growth of Nations

1.1 Eminence of Economic Growth Theories

The quest for finding causal paths and theories for growth, particularly via the lens of economics, has been occupying social scientists since the beginning of the 20^{th} century (DeLong, 2000). As the importance of population and army size gave its place to technology and productivity en route to gaining power in the international arena, governing bodies of nations diverted their focus on creating more wealth, development; and gaining competitive advantage in global markets (Liucvaitiene et al., 2013).

After a myriad of growth theories concocted through various ontological and epistemological perspectives, the theory that served as the basis for future growth theories, and furthered the understanding of how an economy functions was the famous '**Cobb–Douglass Production Function**' (Cobb and Douglas, 1928).As Felipe and Adams (2005) puts it, "the Cobb–Douglas production function is still today the most ubiquitous form in theoretical and empirical analyses of growth and productivity [p.428]." In simple, a production function is a mathematical condition that relates to the maximum amount of output that could be generated from the inputs listed in the function (Jones, 2005). Interested in modeling the GDP of a country into a simple equation, the CDPF is essentially a multiplicative function (Clinton, 2004),

$$CDPF: \qquad Y_t = F(K_t, L_t) = K^{\alpha} L^{1-\alpha} \tag{1.1}$$

where (Douglas, 1976):

- Y: Output (GDP)
- K: Capital (goods such as equipment, dams, machines, and buildings (Samuelson et al., 2011))
- L: Labor
- α : Income Share $\implies$ $0 < \alpha < 1$

In this form, to calculate the additional output created by adding an additional/effective capital, hence the Marginal Product of Capital, partial derivatives of both sides with respect to time, $\frac{\partial Y}{\partial K}$, (Barro et al., 2017)

$$MP_K : \qquad \frac{\partial Y}{\partial K} = \alpha K^{\alpha-1} L^{1-\alpha} \tag{1.2}$$

It is important to notice that $K^{\alpha-1} = K^{\alpha} \cdot K^{-1}$, therefore we can rewrite the equation as

$$MP_K : \qquad \frac{\partial Y}{\partial K} = \alpha \frac{K^{\alpha} L^{1-\alpha}}{K} \tag{1.3}$$

Finally, given the numerator of MP_K equation equals to the output, Y, in equation (1.1), it can be ultimately concluded that:

$$MP_K = \alpha \frac{Y}{K} \tag{1.4}$$

The same mathematical approaches used in equations (1.2:1.4) can also be used to calculate the additional output created by an additional/effective labor(worker), Marginal Product of Labor, by taking partial derivatives, $\frac{\partial Y}{\partial L}$, and then simplifying to get :

$$MP_L = (1 - \alpha) \frac{Y}{L} \tag{1.5}$$

As seen in equations (1.4) and (1.5), both MP_K and MP_L are products of coefficients α and $(1 - \alpha)$, respectively, where the coefficients "are both between zero and unity" (Clinton, 2004, para. 1). Given this, theoretically and practically, the returns generated from adding a unit of effective capital and labor should not sustain a positive increase

in perpetuity given α would always be less than zero (Barro et al., 2017). Henceforward, the returns generated from both MP_K and MP_L, after a certain point of time providing positive returns would give its place to decreasing returns at a certain time point until the economy ultimately reaches to a '*steady state*' level and returns ending up plateaued Solow (1956) (see **Figure 1.1**). This understanding of CDPF, especially the concept of returns flattening at the steady-state level is the main finding of Solow (1956) which is discussed in Section 1.1.1); however, mentioned here to reflect a better understanding of the partial derivations and their interpretations.

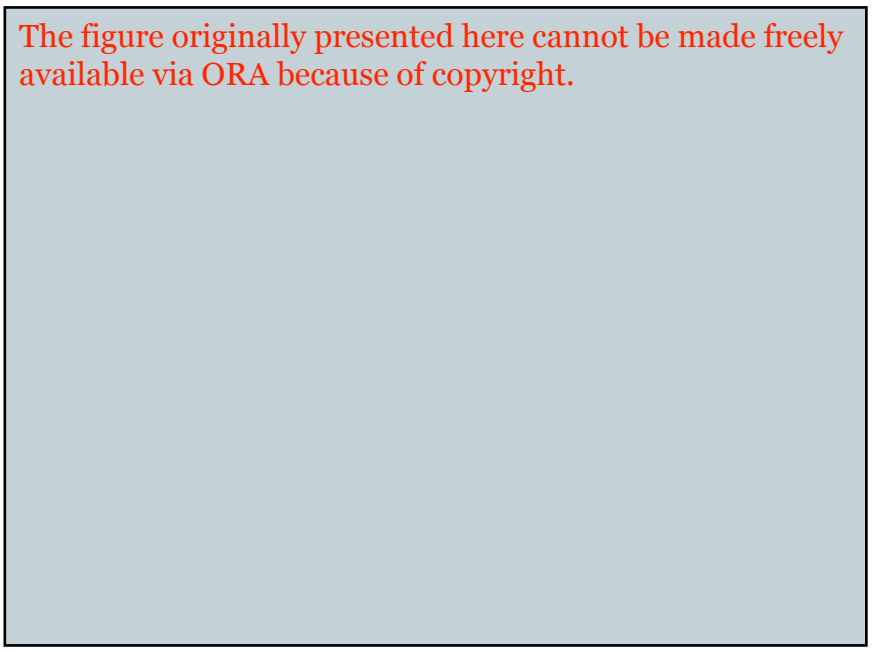


Fig. 1.1 An Example of Global Production Function where Returns from GDP Per Capita and Effective Capital are Increasing at a Decreasing Rate (Jones, 2005, p. 527)

The CDPF was also buttressed by statistical evidence in Cobb and Douglas (1928)'s work where they modeled the growth of the US Economy with data from "US manufacturing sector for 1899-1922" (Felipe and Adams, 2005, p. 428). However, the CDPF was not sufficient to answer greater questions in economics, exempli gratia, other major factors that could have a profound impact on economic growth, capital, and labor. Ultimately, it is vital to understand the dynamics that lie beneath the CDPF– particularly the mathematical proof of diminishing returns from both capital and labor as it led social scientists to ponder upon other variables such as skills formation that could impact economic growth (discussed in Section 1.1.1).

1.1.1 Solow—Swan Model

While the CDPF served as the groundwork where countless scientists relied on to improve the understanding of how an economy grows, the Solow (1956) & Swan (1956) model has become one the most popular theories in economics where scientists still make use of it today. In his paper, Solow (1956) states that:

> All theory depends on assumptions which are not quite true. That's what makes it a theory. The art of successful theorizing is to make the inevitable simplifying assumptions in such a way that the final results are not very sensitive [p.65].

Swan (1956) further argued that given theories are based on '*critical assumptions*,' they are only reliable if the critical assumptions are as realistic as possible. With the intention of creating more realistic models, Solow (1956) & Swan (1956) expanded the CDPF with differential equations type mathematical proofs by breaking down capital, labor and other variables to investigate the mechanisms of economic growth. In the '*Basic Solow–Swan Model with Technological Progress*,' growth is explained as:

$$Y_t = AK_t^{\alpha} L_t^{1-\alpha} \tag{1.6}$$

where 'A' is technological progress or economic growth not explained by capital or labor. (Acemoglu, 2014; Solow, 1956). When partial derivatives with respect to time are taken[1] (Canarella and Pollard, 2011):

- Technology grows in time (t) as:

  $\implies \dot{A}_t = g(\mathrm{A})_t$ where g=growth rate

- Capital grows in time (t)as:

  $\implies \dot{K}_t = s_k \mathrm{Y(t)} - \delta \mathrm{K(t)}$ where s_k = investment in capital, and δ = depreciation

- Labor grows in time (t) as

  $\implies \dot{L}_t = n\mathrm{L(t)}$ where n = population growth

Solow (1956) and Swan (1956)'s work illustrates how concepts such as a country's saving rate, depreciation in both capital and labor, population growth and supply of labor, technological progress, and productivity impact economic growth. In his

[1] The dots on top of variables, e.g. $\dot{A}_t$, indicate partial derivatives with respect to time

paper, Solow (1956) argues that without technological progress, (A),**the maximum level** of economic growth countries could reach would not differentiate among them. Technological progress in the Solow Model is also referred to as the *Total Factor Productivity* (TFP) since it is a measure of economic efficiency and the productivity of people (Romer, 1990). Increases in savings, investments, population growth and other variables except the TFP would only help countries move along the curve of their economic growth (illustrated in **Figure 1.1**), but the maximum growth one country could reach in the long–run would not be any different from the other. However, according to Solow (1956), the only variable that determines both the level and the maximum amount of growth any given country could generate, and create more growth, in the long run, is the level of technological progress of that country, hence TFP.

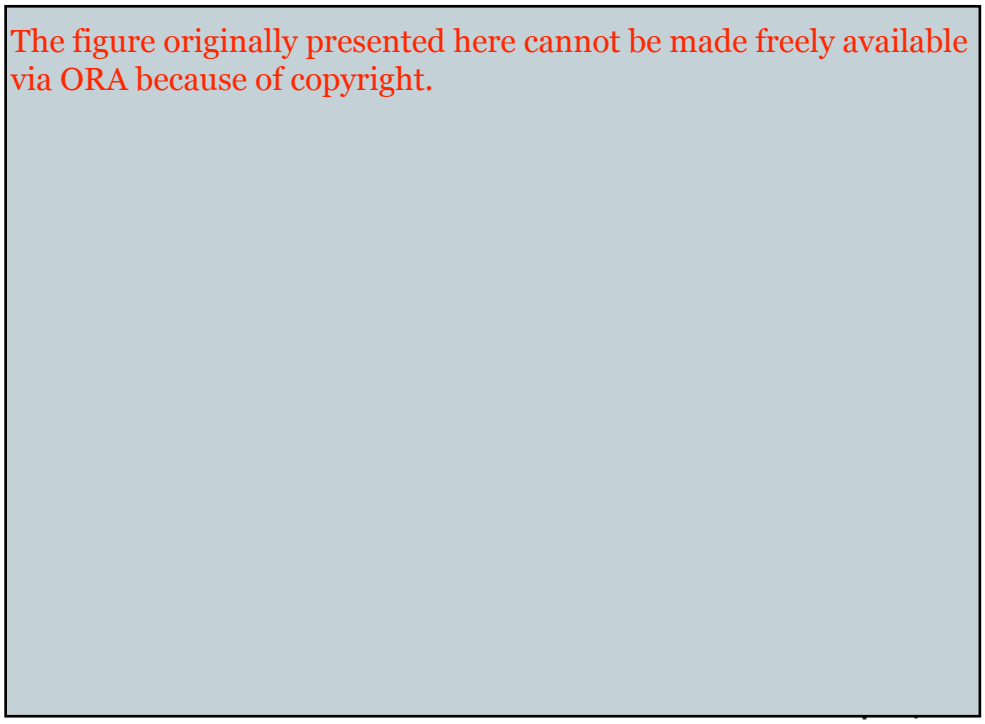


Fig. 1.2 Technological Progress and Economic Growth in the Solow Model. In this example, the increase in capital due to an increase in technological efficiency. (Whelan, 2015, p. 12)
Note: δ_k = Capital Depreciation

Despite producing analogous growth models and explanations, given Solow's work was published 11 months earlier than Swan's led Solow to be awarded the Nobel Prize in 1987 based on his findings on the importance of TFP, which has now entered the literature as '*the Solow Residual*' (The Nobel Prize, 1987). Notwithstanding this fact, both economists are revered as the fathers of Neoclassical Economic growth models (Barro et al., 2017). Solow's finding on how technological progress and the productivity

of people account for cross–country economic growth differences were extremely crucial for the field of development and growth. However, both Solow (1956) and Swan (1956)'s discovery did not delineate what causes an increase in TFP, skills, and productivity. This lack of explanation led scientists to investigate causal paths that would increase skills and productivity (Barro, 2001), and could perhaps be the catalyst that nudged the debate on the importance of education in economic growth into the global arena.

1.2 Human Capital Theory: A Global Phenomenon

With the advent of the Solow–Swan Model, "much attention of macroeconomists has focused on long-term issues, notably the effects of government policies on the long-term rate of economic growth (Barro, 2001, p 1)," given technological progress and skills accumulation take place in the long-run. Henceforward, economists were posed with the challenge of providing theoretical understandings and empirical studies for factors that increase productivity and skills, which, according to Solow (1956), leads to an increase in technological progress. A sizable motivation behind the quest for these factors was that the inexplicable growth from neither capital nor labor, the Solow residual (TFP), accounted for approximately 87.5% of the growth in GDP per capita (GDPPC) in the US between 1909–49 (Acs et al., 2014). Much of that residual was due to the increase in "knowledge creation and the augmentation of the labor input through education and training" (Goldin, 2014, p 2), in other words, '*human capital*' (Hanushek et al., 2011).

Johnes (1993) states that the birth of '*economics of education*' when a Nobel Laureate Economist Theodore Schultz gave a lecture to the American Economic Association titled as "**Investment in Human Capital**(Schultz, 1961). Stating his outright surprise as to why economist waited so long to enter the field of education given how countries allocate massive resources into it, Johnes (1993) defines the economics of education as :

> Economics is the study of choice. Education is the augmentation of stock of skills, knowledge, and understanding possessed either by individuals or by society as a whole. The economics of education, therefore, concerns the manner in which choices affecting this stock are made, both by individuals who demand education and by the teachers and institutions which supply it [p.1].

It was not long enough for the term 'human capital' to percolate within the academic and public policy communities (Becker, 1993) and be picked up by politicians and

country leaders promising investment in human capital (Johnes, 1993). The traces of the concept of human capital, which is the knowledge, skills and the experience of people (OECD, 2009), and investments that increase individual productivity such as education, training, and health (Goldin, 2014) could be found in the father of modern capitalism Adam Smith (1786)'s famous book *The Wealth of Nations*. As per cited in Goldin (2014), Smith (1786) identifies human capital as one of the four definitions of capital by stating :

> The acquisition of such talents, by the maintenance of the acquirer during his education, study, or apprenticeship, always costs a real expense, which is a capital fixed and realized, as it were, in his person. Those talents, as they make a part of his fortune, so do they likewise of that of the society to which he belongs. The improved dexterity of a workman may be considered in the same light as a machine or instrument of trade which facilitates and abridges labor, and which, though it costs a certain expense, repays that expense with a profit [p.123].

1.2.1 Education, Human Capital and Economic Growth

Despite Adam Smith's unwavering statement on the importance of skills of the labor force, human capital became popular and gained prominence after nearly two centuries from Smith's work when (Becker, 1962)'s a study on estimating the impact of educational investments on human capital and economic growth. In this paper, (Becker, 1962) argued that advancements in human capital via investing in education would provide higher yields and rate of returns in juxtaposition to other forms of capital. (Becker, 1962) also stated that much of the unexplained portion of countries' economic growth, the Solow residual, could be explained with investing in education, which he claimed to be an essential form of human capital investment. The reason of this statement mentioned in his later book is because new technological advancements in countries with low levels of skills would be of low value to countries with a modicum amount of people that have the capabilities to make use of them Becker (1993). "Economic growth closely depends on the synergies between new knowledge and human capital, which is why large increases in education and training have accompanied major advances in technological knowledge in all countries that have achieved significant economic growth (Becker, 2002, para 18)" Henceforward, Becker (2002) explains the considerable variance in GDPPC across countries with investment in education, which translates into skills that could produce and operate new technologies.

After Becker's work, other researchers devised alternative methods to explain the unexplained growth in the Solow Model by incorporating variables that would account for human capital (Goldin, 2014). In their work, by stating "This paper takes Robert Solow seriously [p. 407]" Mankiw et al. (1992) used regression analysis in the form of panel data and created the '*Technology Augmented Solow-Swan Model with Human Capital*' :

$$Y_j(t) = K_j(t)^{\alpha} H_j(t)^{\beta} (A_j(t) L_j(t))^{1-\alpha-\beta} \tag{1.7}$$

where:

- j: Country j=1...N has aggregate CDPF (Acemoglu, 2014)
- H: Human Capital

  $\Longrightarrow$ "The amount of human capital is H(t) $\equiv$ L(t) G(E) [own emphasis added], which is the number of workers L(t) times the amount of human capital per worker G(E) , where E is the average education level of current workers [p.58]" (Parker, 2015).

- α and β : Income Share

  $\Longrightarrow$ $0 < \alpha < 1$

  $\Longrightarrow$ $0 < \beta < 1$

  $\Longrightarrow$ $\alpha + \beta = 1$

In Mankiw et al. (1992) 's augmented Solow Model, while the assumptions for how capital and labor grow in time are the same with the Basic Solow Model in equation (1.6), human capital in time (t) grows as : $\dot{H}_t = s_h \mathrm{Y(t)} - \delta \mathrm{H(t)}$ where 's_k' is the investment in human capital, and 'δ' is the depreciation of human capital (Canarella and Pollard, 2011). It is important to note that Mankiw et al. (1992) 's model, following Smith (1786)'s definition of human capital, treats the human capital variable (H) in their equation no different than physical capital (K), where growth in time, $\dot{H}_t$, decreases with depreciation in human capital. This depreciation could be articulated as the decrease in the economic value skills in horse riding or swordsmanship from 1800-2019. Besides, the direct impact of education on human capital is shown in Mankiw et al. (1992)'s augmented Solow Model as the average education level is a factor of human capital. It is noteworthy to mention that the technology augmented Solow Model with

the incorporation of human capital version also served as the **theoretical framework inspiration** for the statistical analyses in the further chapters of this thesis.

Mankiw et al. (1992) 's regression models using equation (1.7) as a framework were able to explain roughly 80 % of the variation in cross-country GDP growth rates — a stark increase in comparison to the previous models such as Cobb and Douglas (1928); Solow (1956); Swan (1956); and Becker (1962). The data was split into three samples: countries where oil production is not the main source of income; countries with a population of less than a million; and finally 22 OECD countries. Although human capital encompasses a range of numerous variables that impact human productivity, only educational attainment, and investment variables were used in Mankiw et al. (1992)'s analyses. The authors note that this posed a series of problems. To begin, Mankiw et al. (1992) state that the aggregate/lump-sum value of the educational investment is problematic given that not all investment in education provides the same rate of return and merely spending more on education does not translate into higher returns or productive human capital. "Philosophy, religion, and literature, for example, although serving in part to train the mind, might also be a form of consumption (Mankiw et al., 1992, p 419). Furthermore, the authors only use secondary educational attainment figures, which is itself far from being axiomatic given it does not capture the true effect education could have on economic growth. While the authors do not shy away from accepting the problems associated with using secondary level school attainment variables as they state "despite this narrowed focus, measurement of human capital presents great practical difficulties Mankiw et al. (1992, p 418)," the vociferous debate as to whether the application of human capital theory could account for the value provided by education had already begun, predominantly from voices outside of economics (Tan, 2014).

1.3 Rationalizing the Impact of Education on Economic Growth

Acs et al. (2014) mention that just as "the Higgs boson, an elementary building block of modern physics, was first conceptualized in 1964 and its existence only confirmed in 2013 [p.2]," the inability to shed light on the Solow residual–similar to the Higgs boson–has taken up dire efforts of economists and social scientist. Although the inclusion of human capital (mainly educational variables) into economic growth functions was helpful in terms of elucidating other aspects of economic growth, the variable brought on a myriad of problems along with itself (Barro and Sala-i Martin, 2004).

Many influential publications in the field of economics of education such as Barro and Sala-i Martin (2004); Hanushek and Woessmann (2015); Romer (1986) and countless others have all extensively mentioned various problems associated with the usage of school attainment figures to account for the impact of education on economic growth. In the book of the creator of Barro and Lee (2013) global educational attainment data set, Robert Barro, and economist Sala-i Martin, the authors widely discuss the problems they encountered with educational variables (Barro and Sala-i Martin, 2004). In their three-staged ordinary least square models[2], Barro and Sala-i Martin (2004) attempt to explain the cross-country growth rate of GDPPC where they divide the countries into two samples of low and high based on their current level of GDP. The authors use variables such as life expectancy, fertility rate, the rule of law, democracy indicator, openness ratio to account for how exposed the country is to international trade, and other variables in the Solow Model. Some of the commonalities among prosperous countries were high levels of educational attainment, saving rates, openness in international trade, and vigorous enforcement of the rule of law through democratic political institutions. When it comes to the educational variables; however, only the **male upper-level schooling** variable is used in the regression models. The authors state the reason for that as :

> Female upper-level schooling has a negative but statistically insignificant coefficient, - 0 .0034. Schooling at the primary level for males or females also has statistically insignificant coefficients: - 0 .0011 and 0.0007, respectively. Hence, the main relation between growth and years of schooling involves the male upper-level component[own emphasis added]. A separation of this male variable into college and high-school components generates two positive coefficients, 0.0105 and 0.0024, that are insignificantly different from each other [p.537].

Given the human capital theory is criticized for not taking outside factors such as the political situation in countries into account (Tan, 2014), the inclusion of such factors in Barro and Sala-i Martin (2004)'s models were conducive to assuage some of the sharp criticism on using empirical models to test out the impact of education on economic growth. Although the qualitative variables in Barro and Sala-i Martin (2004)'s model improve the models' explanation of the dependent variable, the problems associated with "the quantity of education, as measured by years of schooling, rather than the quality"(Barro and Sala-i Martin, 2004, p 537) is evident. The authors suggest

[2]The results from their OLS models are shown in Appendix A

that internationally comparable student outcomes scores could be used as means to the quality of education, which the results from them may elicit factors outside of formal education such as *the family's influence on student achievement*. This suggestion is crucial in terms of serving as an explanation of the analyses that will be conducted in the subsequent chapters of this thesis.

Barro and Sala-i Martin (2004) were only a few of the many social scientists that encountered significant problems using the quantity of education in their models. For instance, in their analysis, Mankiw et al. (1992) state that while their technology augmented Solow Model with Human capital explained the variation in GDPPC growth of the majority of the countries in the world, it had low explanatory power for their OECD sample of at the time 22 countries (Canarella and Pollard, 2011). However, Barro and Sala-i Martin (2004)'s suggestion of using measures of quality of education in lieu of quantity had already been initiated by Hanushek and Kimko (2000), which is discussed in the following section, **1.3.1**.

1.3.1 The Role of Cognitive Skills in Educational Quality

One of the most omnipresent forms of criticisms on human capital is on how the widely-claimed impact of human capital accumulation (using educational attainment figures) was unable to predict the growth in GDPPC across the world (Marginson, 2017). Marginson (2017) state that although the human capital theory has dominated economics and public policymaking since the 60s, it has failed to produce realistic results. As mentioned in (Tan, 2014), Pritchett (2009) show that average school attainment showed either nonlinear or negative relationship with economic growth in Latin American countries such as Bolivia, El Salvador, Mexico, and, Peru. Furthermore, Hausmann and Rodríguez (2014) show in Figure **1.3** average years of total schooling show a complete separation and a negative relationship with GDPPC starting 1975.

As the criticisms for theories suggesting human capital formation as a critical component for economic growth were inundating social science journals, Hanushek and Kimko (2000)'s quality over quantity approach saved human capital from a derailment. In their research, concerning the previous approaches used to measure human capital, Hanushek and Kimko (2000) state that:

> Virtually all [human capital models] have ignored quality issues, implicitly assuming that any variations in the quality of human capital are small relative to the importance and variation in pure quantity of human capital [p. 1203].

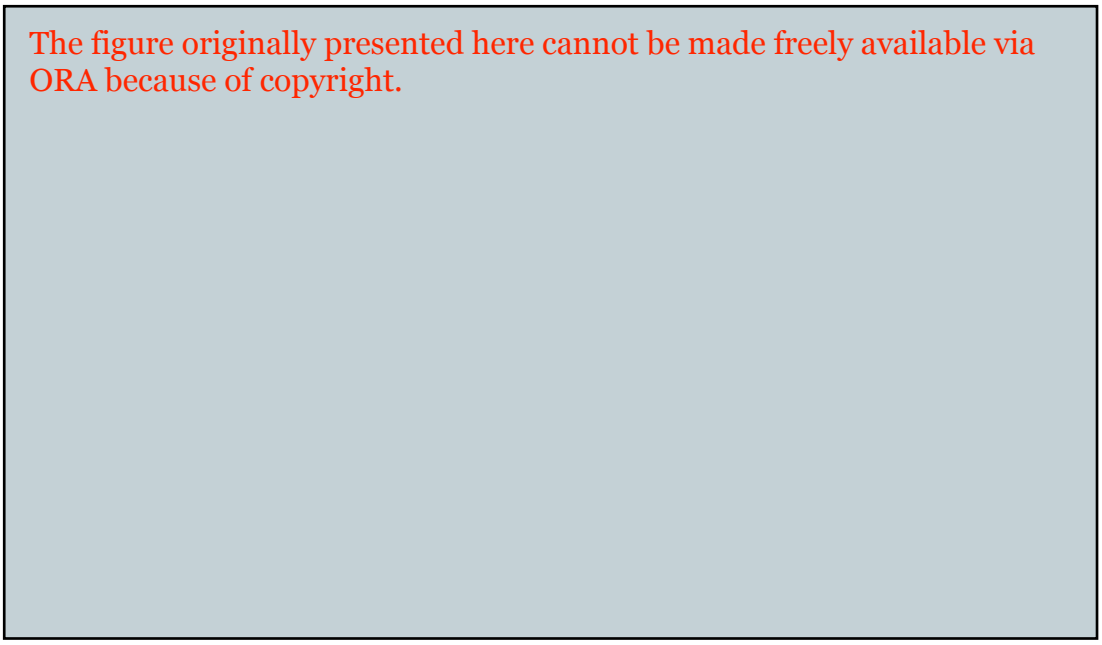


Fig. 1.3 Educational Attainment and GDPPC (PPP) in Venezuela, 1960 – 2000 **Source:** Hausmann and Rodríguez (2014) (Also cited in Tan (2014, p 425))

Hanushek and Kimko (2000)'s research brought a profound impact into the economics of education where the authors show how concepts such as '*quality of labor force*' and '*quality of education*' form a significant relationship with economic growth. Hanushek and Kimko (2000) used different math and science international examination scores produced by the IEA, which were later combined under the widely–known TIMMS, to measure the quality of labor. Some of these famous international examinations, which are still conducted nowadays, are explained in Table **A** in Appendix **A**. Hanushek and Kimko (2000) find that "the growth relationship does not appear to the result of growth causing higher quality through investing resources in schools" [p.1203]. Furthermore, they state that " the estimated impact of quality on growth, indicating that one standard deviation [increase] in **mathematics and science skills** translates into more than one percentage point in average annual real growth, also looks implausibly large." In his later analysis, Barro (2001) state that his findings also corroborate the ones from Hanushek and Kimko (2000) as he, contrary to previous studies (Barro, 1991), also finds that "find that scores on international examinations — indicators of the quality of schooling capital — matter more than years of attainment for subsequent economic growth" [p.23]. These are some of the essential explanations which the statistical investigations conducted in subsequent sections rely on.

American economist Eric Hanushek complied all his years worth of research in 2014 when he published his book '*The Knowledge Capital of Nations*' with German

economist Ludger Woessmann. Both Hanushek and Woessman are ranked within the top 5 economists in the field of education as of May 2019 (IDEAS REPEC, 2019), and their data-driven work has certainly shed more light on concepts such as what aspects of teachers count towards their quality; and what sort of educational investments create economic returns; and how countries should shape educational systems. (Barro, 2001; Hanushek and Kimko, 2000; Hanushek et al., 2011; Hanushek and Woessmann, 2015). It seems that by referring to the education portion of the human capital as *the Knowledge Capital*, Hanushek and Woessmann (2015) make it clear that human capital, as mentioned previously in **1.2.1**, encompasses all possible variables that impact human productivity. This distinction from human to knowledge capital also addresses some criticism on how only using education as a measure of human capital oversimplifies the concept (Marginson, 2017).

Hanushek and Woessmann (2015) findings also suggest that the relationship between merely keeping children in school and lengthening years of education are inconsistent and insignificant with economic growth. From a theoretical standpoint, this finding also makes intuitive sense. For instance if, ceteris paribus [3], the quality of education provided to students is scant, neither the number of people in school and nor their average years of education would not be expected to contribute both to the person and on a macro-level to the country. This statistical support for these claims is discussed in section **1.3.2**.

1.3.2 Solving the World's Growth Paradox Through Cognitive Skills

Hanushek and Woessmann (2015) start their book with heavy criticism on previous human capital measurement methods. They state that the world's attention, driven by governments and international organizations like the World Bank and the UN, has focused on quantity of education such as school inputs (spending, class size, etc.) and issues regarding quality were ignored. Policies constituted around economic growth that followed the quantity approach to form human capital have failed, which created disappointment and strong disapproval. Therefore, "the significance of education [own emphasis added] has been obscured by measurement issues" (Hanushek and Woessmann, 2015, p 2). Furthermore, the authors state that the ultimate conclusion of their book is compatible with Smith (1786)'s statement on the importance of human capital in section **1.2**; however, it has not been appropriately measured. Notwithstanding the

[3] All other things are equal/held constant

fact of being what academics might call 'a powerful statement' the pith of Hanushek and Woessmann (2015) argument is:

> Although many factors enter into a nation's economic growth, we conclude that the cognitive skills of the population are the most essential to long-run prosperity. [Own emphasis added]. . . the "knowledge capital" of a nation, explain in large part the differences in long-run growth we have seen around the world in the past half-century. Moreover, in the presence of measures of cognitive skills, school attainment does not even have a significant independent relationship with growth. [p.2]

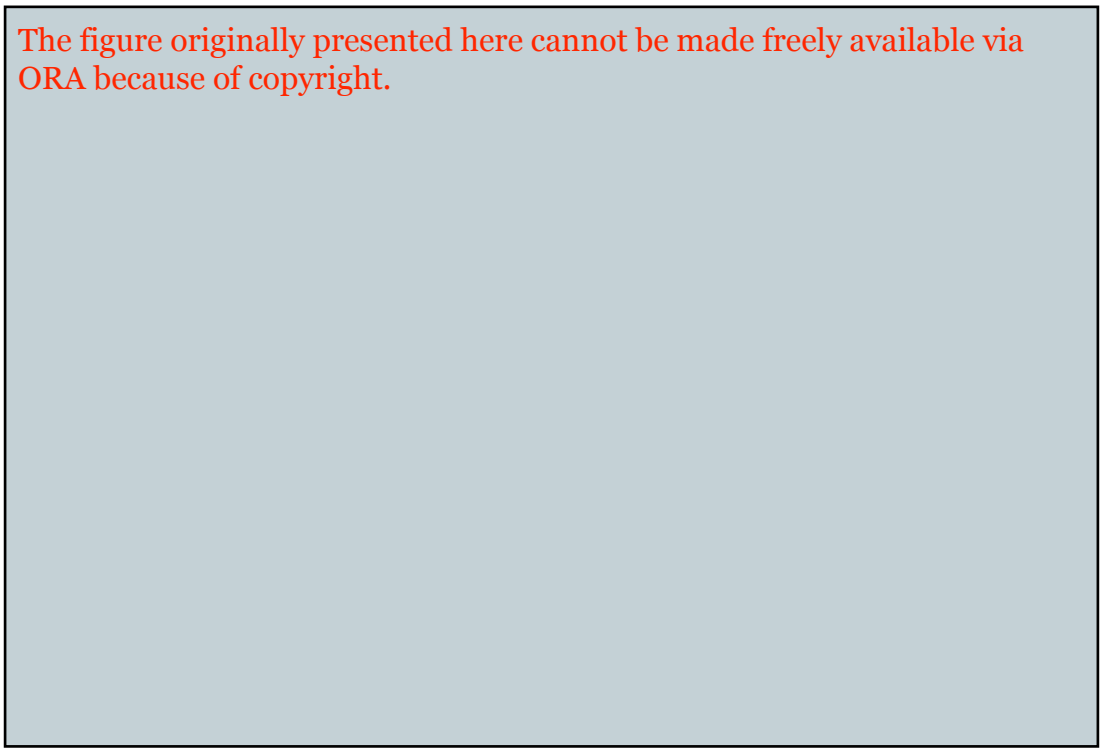


Fig. 1.4 Knowledge Capital and Economic Growth in the World. Plot shows a regression of GDPPC growth in 1960-2009 on average international student achievement tests. (Hanushek and Woessmann, 2015, p. 4) **Note:** See Appendix **A** for the region abbreviations.

The analysis in Hanushek and Woessmann (2015) starts with untangling the growth paradox in the world that previous studies were unable to explain using school attainment as an educational measure. The authors state that in the 1960s, GDPPC in Latin America was higher than sub-Saharan and MENA [4] regions, and

[4] Middle East and North Africa

both Latin American and African countries had higher GDPPC than East Asian countries. Subsequently, Latin American countries had higher school attainment figures than MENA and East Asia regions; therefore, one might expect to Latin America to take off in economic growth by observing these human capital investments. However, "today, East Asia has moved far ahead of Latin America in growth and income. The MENA region has also jumped ahead [emphasis added], leaving only Latin America and sub-Saharan Africa at the bottom with very low long-term growth rates and commensurate low income per capita"(Hanushek and Woessmann, 2015, p 3). When cognitive skills are plotted against the conditional GDPPC growth rate (see Figure **1.4**), the answer for the growth paradox seems to unravel.

Cognitive skills, which include literacy and numeracy skills, are defined as the "ability to understand complex ideas, adapt effectively to the environment, learn from experience, and reason" (World Bank, 2017b, para 5). UNESCO (2004) describes the importance of cognitive development as "the major explicit objective of all education systems" [p.2] and as one of the principles of defining quality in education. In Hanushek and Woessmann (2015), cognitive skills are referred to as Knowledge Capital, and the authors believe that **international mathematics and science** scores are adequate measures for cognitive skills. The predictive power of these math and science scores in juxtaposition to educational attainment can be seen in Table **1.1**. While educational attainment can only predict 25% of the change in the annual rate of GDPPC growth in 50 countries from 1964 to 2003, international tests in math and science (cognitive skills) alone predict around 73% of the variation with the absence of the schooling variable. Educational attainment seems to show an insignificant statistical relationship with economic growth when the variable that accounts for cognitive skills is entered into the models; hence confirming Hanushek and Woessmann (2015)'s quote in section **1.3.2** [para 1].

It should be noted that although cognitive skills have significant explanatory power in per-capita GDP growth in the world, it is not the only variable that impacts how countries grow– neither Hanushek and Woessmann (2015) state such. There are other factors as necessary, if not even more important to also have the basis to provide quality education within a country.

Table 1.1 Hanushek and Woessmann (2015)'s OLS Models

	Dependent variable:						
	Annual Rate of GDPPC Growth						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cognitive Skills		2.015***	1.980***	1.975***	1.933***	1.666***	1.985***
		(10.68)	(9.12)	(8.28)	(8.29)	(5.09)	(7.83)
Initial years of schooling (1960)	0.369***		0.026		0.025	0.047	-0.090
	(3.23)		(0.34)		(0.29)	(0.54)	(1.02)
Initial GDPPC (1960)	-0.379***	-0.287***	-0.302***	-0.298***	-0.298***	-0.255***	
	(4.24)	(9.15)	(5.54)	(6.02)	(5.04)	(3.12)	
Average years of schooling (1960,2000)				0.024			
				(0.78)			
log (initial GDPPC (2000)							-0.879***
							(3.39)
Indicators for 8 world regions					yes		
Constant	2.785***	-4.827***	-4.737***	-4.764***	-4.536***	-3.185***	-4.491***
	(7.41)	(6.00)	(5.54)	(5.66)	(4.97)	(2.16)	(4.48)
Observations	50	50	50	50	52	50	50
Adjusted R^2	0.252	0.733	0.728	0.728		0.706	0.637

Notes: Cognitive skill measure refers to average score on all international tests 1964 to 2003 in **math and science**, primary through end of secondary school. *t*- statistics in parentheses; statistical significance at *p<0.1; **p<0.05; ***p<0.01
Source: (Hanushek and Woessmann, 2015, p 43)

The plight of Latin American countries is a topic highly inquired within the fields in social sciences (Hanushek and Woessmann, 2012). One of the most influential books in economic development through the lens of history and political science is *Why Nations Fail* by Acemoglu and Robinson (2012). One of the examples in their extensive historical and economic analysis starting from the creation of first civilizations is a town called Nogales. Nogales is a shared territory, separated with nothing but a fence, that lies between the US and Mexico. While Nogales, Arizona has high levels of income per person, unwavering infrastructure, reliable health care services; Nogales, Sonora in Mexico has on average $1/8^th$ of the income in the US side of the border, and lest, battles with significant corruption and crime issues. Acemoglu (2014) reason that this issue is due to the different type of institutions that govern this territory. The US has, what the authors call 'inclusive institutions,' which they define as "those that allow and encourage participation by the great mass of people in economic activities that make the best use of their talents and skills and that enable individuals to make the choices they wish" (Acemoglu, 2014, p 74). However, the institutions in countries like Mexico are classified as 'extractive institutions,' "because such institutions are designed to extract incomes and wealth from one subset of society to benefit a different subset" (Acemoglu, 2014, p 76).

Both claims in Hanushek and Woessmann (2015) where low economic performance is identified with commensurately low levels of cognitive skills accumulation, and in Acemoglu (2014) with inclusive vs. extractive institutions seem to hold ground when the sources of economic growth are analyzed. Building on the US vs. Mexico example in Acemoglu (2014), the growth rates of these countries can be analyzed to see how much contribution comes from TFP, labor, capital, and human capital using Mankiw et al. (1992)'s Human Capital Augmented Solow Model and its derivations in section **1.2.1**. The data came from Penn World Tables (PWT) 9.0 (Feenstra et al., 2015; Zeileis, 2017), and only the years between 1950-2014 to test Hanushek and Woessmann (2015), and Acemoglu (2014)'s statements within the respected years of their publications. As per seen in Table **1.2**, a significant share of economic growth in the US (40.16%) was generated through human capital and TFP; while a striking 83.90% of the growth came from labor and capital in Mexico. Henceforward, differences in the sources of economic growth provide additional reasoning on the stark contrast between these countries' income levels, living conditions, innovation creation shares, and educational quality. Even though Mexico's economy grew around 1.21% more on average per annum, it is still behind the US in almost every indicator of economic and human well-being (OECD, 2018). From thereof, countries that do not base their growth trajectories on

the cognitive skills accumulation and inclusive institution creation seem to fall behind of countries that do as seen in cross-country analyses of Hanushek and Woessmann (2015) and Acemoglu (2014) and results shown in Table **1.2**.

Ultimately, it could be argued that the firm claims starting in Hanushek and Kimko (2000) and later in Hanushek and Woessmann (2015) on the importance of international science and mathematics scores nudged significant initiatives taken by national governments and international organizations. The PISA nowadays is a global phenomenon which created unheard stars in quality of education such as Finland (Choi, 2015); and also gave comparative analyses to parents of 15-year-olds by exposing countries where educational quality is meager. On the other hand, the World Bank and the UN have prioritized quality education by making it one of the 17 Sustainable Development Goals in 2030, and human capital formation an vital indicator of development (Sustainable Development Goals, 2015; World Bank, 2017a).

Table 1.2 Economic Growth Accounting for Mexico and the US between 1950-2014 using the Human Capital Augmented Solow Model

	Mexico	**US**
GDP Growth Rate	4.321	3.106
TFP Contribution	0.273	0.903
Labor Input Contribution	1.378	0.729
Human Capital Contribution	0.423	0.358
Physical Capital Contribution	2.249	1.125
Growth from Labor and Physical Capital	**83.90**	59.37
Growth from Human Capital and TFP	16.10	**40.63**

Notes: All numbers are percentages. Growth accounting was done using the Human Capital Augmented Solow Model and its derivations in Mankiw et al. (1992).

Data: Penn World Tables 9.1 (Feenstra et al., 2015; Zeileis, 2017)

To conclude on the importance of cognitive skills, Hanushek and Woessmann (2015)' final suggestions are what helped shaped the support for 1) the vital importance of mathematics and science skills, and 2) why this thesis is concerned with investigating cognitive skills in Sudan. The following findings are in Hanushek and Woessmann (2015, p 203).

- In the long run, a nation's prosperity is directly related to the skills of its population.
- The relevant cognitive skills — what we call the knowledge capital of a nation — can be measured well by performance on **international math and science assessments**
- As demonstrated by historical experience, relatively modest improvements in skills can be worth multiples of a country ' s current GDP.
- Reform is possible even if difficult. The most productive reform involves aligning incentives with achievement through better educational institutions.
- Nations that successfully reform will systematically pull away in economic terms from those that do not. In short, to us the evidence points unwaveringly toward a unified theory of **long-run growth based on the knowledge capital of nations.**

1.4 The Impact of SES on Cognitive Skills

The importance of cognitive skills for long-term growth is well-established with no dearth of empirical support (Barro, 2001; Hanushek and Kimko, 2000; Hanushek and Woessmann, 2012, 2015; OECD, 2009, 2018; World Bank, 2017a,b). Therefore it would be wise for countries seeking development to figure out certain elements and causal paths that could profoundly impact cognitive skills development. There are a great number of factors impact one's cognitive skills formation: parental genetic influences, socioeconomic (SES) status (Figlio et al., 2017), and investment in their children's skills (Cunha and Heckman, 2007, 2008) ; education systems and schools (OECD, 2016); environmental factors (Heckman and Mosso, 2014); non-cognitive factors like perseverance, resilience (Cassen et al., 2009), and grit (Reed and Jeremiah, 2017); and possibly other factors given how complex and yet to be fully mapped out the brain is (National Institutes of Health, 2018).

The reason this thesis concerns predicting the impact of socioeconomic factors on student outcomes in mathematics and science (given the importance of these skills established in (Hanushek and Woessmann, 2015)) in Sudan is a three-fold. First, the data set from the 2018 Sudanese National Learning Assessment provides a survey on the SES of the parents, such as their education status and types of profession. Furthermore, the students were surveyed on whether they had access to certain goods. Therefore, the best fit for the available data is to conduct analyses to predict the impact of socioeconomic factors on science performance.

Second, Boo (2014) mentions the fact on how research shows that "poor nutritional status and low levels of cognitive development in early childhood are important determinants of outcomes such as fewer years of schooling, less learning while in school, lower cognitive skills and worse health in adolescence (including mental health) and adulthood, a lower probability of employment, lower earnings, lower wage rates, and more criminal activity" [p.1]. Alternatively, studies such as Cunha and Heckman (2007) where the authors state that "skills beget skills and abilities beget abilities" [p.8], theorize that the environment which a child grows up, along with investment in his/her skills are two of the three predictors of the skills[5] that could possibly be formed in the future, which both are significantly depended on parental SES.

Although the theory is clear, findings on the impact of parental SES on student outcomes alter with a high magnitude across countries Boo (2014). Per exempli, while Abuya et al. (2013) find that mother's education predicted girls' learning outcomes in Kenya; Osullivan et al. (2005) elsewhere find that maternal education had a stronger effect on boys' achievement. Furthermore, in her analysis, Boo (2014) uses an extensive household survey provided by Young Lives and finds that the impact of SES was the highest for the most unequal country in her country sample, which was Peru [6], for predicting "receptive language ability among children aged 55 to 102 months" [p.2]. This finding is also supported in other studies, such as Hertz et al. (2008), where the authors find that the impact of the SES on intergenerational persistence of educational attainment was the highest in Peru in comparison to the countries used in Boo (2014)'s sample. Finally, it seems that the impact of SES on cognitive skills lowers in countries with high levels of the rule of law, openness to trade, and GDPPC (Hanushek and Woessmann, 2012; Heckman and Mosso, 2014). One of the main reasons could be the fact that educational opportunities are more accessible in juxtaposition to low performing countries, and studies such as Cassen et al. (2009), and Reed and Jeremiah

[5]The third predictor is the child's initial level of skills.

[6]The other countries in Boo (2014)'s sample were Ethiopia, India, and Vietnam

(2017) show that in fair conditions, adversity factors such as low parental SES could give rise to components that predict future success such as grit and resilience.

Finally, Boo (2014) notes that "not much research exists on differences in cognitive development by SES, when they arise, whom they affect, and how they evolve as children grow up"[p.1]. This could be because data for parental SES is usually collected through household surveys. Given that:

1. SES data is mostly an ordinal variable collected with Likert scales responses or class ranks, hence the range of the variable is short (Muijs, 2011), and

2. Categorical variables usually have lower and more varying predictive power in comparison to continuous variables (Field, 2018),

researchers might have taken their focus elsewhere. Therefore, to both fill this gap in research and find how parental SES affects mathematics and science outcomes, hence cognitive skills (Hanushek and Woessmann, 2015) in Sudan, this analysis would be beneficial for the scientific community for educational research given the paucity of such research in Sudan (World Bank, 2012).

1.5 Criticism of Human Capital & Measurement Gaps

Almost all of the aggregated functions of economic growth using human capital as a proxy has been receiving incessant criticism across social science disciplines (Tan, 2014). For instance, Becker (1993) reveals 30 years after his famous study Becker (1962), where he argued that advancements in human capital with educational investments would provide higher yields and rate of returns, that he deliberately hesitated to use the term "human capital." Becker (1993) writes that the reason for that is because the human capital approach to vitalize an economy would be considered as if he treated people like slaves. The author states that:

> I hesitated a while before deciding to use the title *Human Capital* for the book I published 30 years ago [emphasis added]. In the early days, many people were criticizing this term and the underlying analysis because they believed it treated people like slaves or machines. [p.85].

This line of criticism on human capital is still prevalent. Although one of the authors of the '*the Knowledge Capital of Nations*,' German economist Ludger Woessmann is

ranked at the fourth-best economist of education in the world (IDEAS REPEC, 2019), ironically, the phrase of 'human capital' was selected as the *Non-Word* of the year in 2011 in Germany. The linguistics experts in the German weekly magazine *Der Spiegel* reasoned this as the term allegedly was inhumanely treating people as mere production factors. There is also growing opposition from business communities as Cohn (2017) argues that the human capital approach merely focuses on cognitive skills while ignoring the other parts of the brain that make up the affective and conative, hence the noncognitive skills.

Although it could be attributed to prior research using educational attainment as a proxy to measure human capital, the criticism by Cohn (2017) seems antedated given the latest developments in research. Many studies state noncognitive skills are also essential factors for economic growth; however, more cross-country research needs to be conducted as current studies can only rely on survey data Garcia (2014); Hanushek et al. (2018); Hanushek and Woessmann (2012), hence unable to draw macro-level inferences. Furthermore, companies such as the Ernst & Young, which referred to as one of the 'Big Four' of the global auditors, dropped their university requirement and now instead tests potential employees with EQ and noncognitive skills assessments Sherriff (2015). Therefore, it seems that both research and some segments of the global business have already addressed Cohn (2017)'s criticism.

One could perhaps write a thesis merely critiquing human capital theories since it is prone to various criticisms like the many theories within the social sciences. Especially the usage of international assessments to measure cognitive abilities is under severe criticism; from how varying examinee initiatives causes skew in results (Gneezy et al., 2017) to how countries try to inflate their results by excluding low-performing students out of their samples (Gruber, 2017). Furthermore, as mentioned in Section **1.3.1**, the most uniform criticism on human capital is on its claimed low predicting impact on GDPPC growth. This is also accepted by Hanushek and Woessmann (2015) and studies in Sections **1.3.2**–**1.4**. What instead should have been used to measure human capital– international mathematics and science assessments hence cognitive skills– are discussed in depth.

Overall, there seems to be a significant misconception between the claimed and perceived effect of human capital on development. It should be observed that the human capital approach to development is one of the many factors of production. From its abstract form in Mankiw et al. (1992), it is treated as a depreciating asset given that skills of yesterday bring lesser output today. The theory also accounts for concepts such as over-qualification as it provides decreasing returns to scale in the long-run.

It is hard to imagine that economists of education are claiming no input other than human capital could affect people's lives and bring prosperity.

Despite many valid critiques, both data and historical accounts show that nations that relied on

1. constituting inclusive institutions to account for the rule of law; protection or private property; openness to trade and other common elements seen in highly democratized governments (Acemoglu and Robinson, 2012), and

2. increasing their citizens knowledge capital which is measured by international math and science scores (Hanushek and Woessmann, 2015)

experienced high levels of prosperity and human development. Overall, as Hanushek and Woessmann (2015) states: "the evidence points unwaveringly toward a unified theory of **long-run growth based on the knowledge capital of nations**"[p. 203].

Chapter 2

Political Economy of Education in Sudan

2.1 Context of the Country

The Republic of Sudan is the third-largest country in Africa, located in the Northeast (Fincham, 2018) (see Fig 2.1) with a size of roughly the half of the EU, around 1.9 million km^2(Central Intelligence Agency, 2019). Sudan borders with countries facing significant fragility and instability (EQUIP-Sudan, 2016) such as Central African Republic, Chad, Egypt, Eritrea, Ethiopia, Libya, and South Sudan(Fincham, 2018). Currently with a population around 41 Million that has been growing at a rate of 2.6 % over the past two decades (World Bank, 2019b), what makes Sudan important for economists of education is that almost 40% of the population is under the age of 14 (EQUIP-Sudan, 2016).

The region where Sudan resides was under Ottoman Empire rule from 1821-1899; then controlled by the British Mandate until 1956 (Fanack, 2018) when Sudan claimed independence in 1956 (Johnson, 2017). Despite gaining its independence, incessant internal conflict and tension between ethnic and religious groups have precluded Sudan's growth (see Fig 2.1) (Fincham, 2018). Sharia Law was introduced and enacted by President Numeiri in 1983 (BBC News, 2019a), which led to decades of conflict between the Christian majority South and the Arab Muslim rule in the North (Fincham, 2018). "Two rounds of north-south civil war cost the lives of 1.5 million people, and a continuing conflict in the western region of Darfur has driven two million people from their homes and killed more than 200,000" (BBC News, 2019a, para 4).

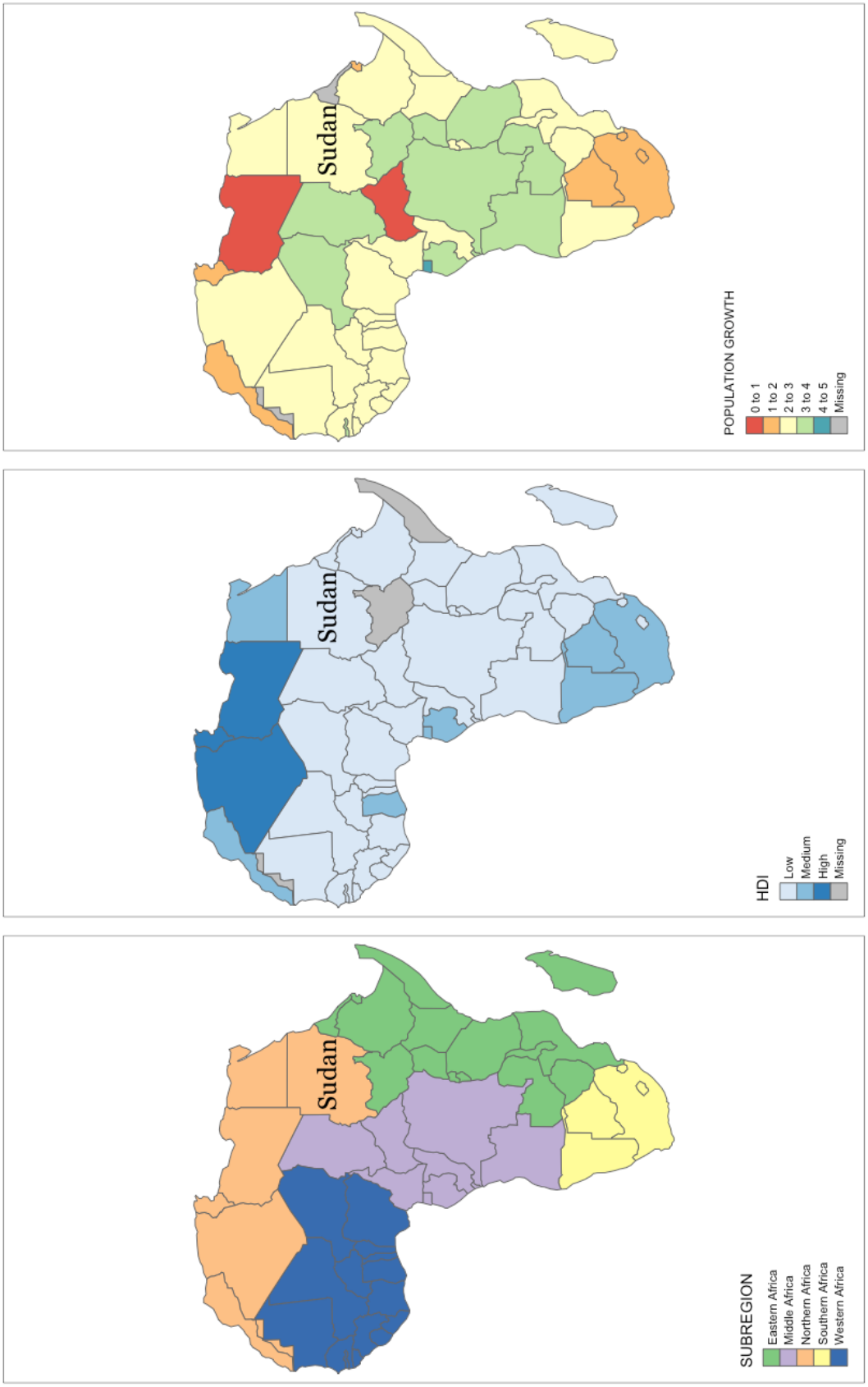


Fig. 2.1 Sub-regions, Human Development and Population Growth in Africa. Data: World Bank (2019b).
Note: See Appendix B for plot creation

As a result of these atrocious events, the International Criminal Court issued two arrests ordered for Sudan's 7^{th} President Omar Al-Bashir in 2009&2010 for the charges[1] of crimes against humanity, war crimes, and genocide. Another reason for this prolonged conflict was the secession of the southern region of Sudan (Fincham, 2018). Almost 99% of the region voted for independence in the referendum, and the Republic of South Sudan became the latest recognized country in the world (BBC News, 2011). The area where South Sudan now occupies held 80% of the oil fields of Sudan – the primary source of Sudanese national income –, which drew the country into further economic stagnation (Fincham, 2018).

The future trajectory of Sudan seems bleak as Al-Bashir's 30-year rule was ended with a military coup d'état on April 11^{th}, 2019 following months of protests due to meager economic conditions and highly inflated cost of living (Deutsche Welle, 2019). Since then, the Sudanese military junta called the Transitional Military Council (TMC) has been in charge despite promising a civilian government (Deutsche Welle, 2019). Even though the talks between the civilian leaders and TMC were initially promising, they reached an impasse on the number of military officials in the transition government (Al Jazeera, 2019). After weeks of negotiation, the TMC ended the talks and brutally cracked down on protesters, which resulted in casualties of more than 112 civilians; more than 70 cases of rape and 800+ injured (Patta, 2019). As a result, the African Union suspended Sudan's membership, and international organizations such as the UN have unequivocally condemned the crackdown on protesters(Burke, 2019).

As of July 5^{th}, 2019, the TMC and the civilians have reached an agreement to share power until elections in 39 months (Walsh, 2019). The transition government will be led for the first 21 months by the military, followed by 18 months of rule by the civilians (Walsh, 2019). The transition government will consist of 11 members, five members from both parties, with the 11^{th} member selected conjointly (BBC News, 2019b).

2.2 The Economic Performance of Sudan

Sudan was the 17^{th} fastest growing economy in 2010 predominantly from oil revenues, and GDPPC growth was around a fourfold between 1996-2001 (UNESCO, 2018). As seen in figure **2.2**, Sudan's GDPPC growth sharply declined following South Sudan's secession (World Bank, 2019b), despite reaching at the peak of 11.83% before the referendum. Since oil exportation accounted for 75% of government revenues and 95%

[1]These charges are in Appendix **A**

of exports for Sudan in 2011, this sharp decline in GDPPC growth is reasoned with the fact that the territory which now South Sudan occupies held more than 80% of Sudan's oil fields (African Development Bank, 2019). Sudan's heavy dependence on oil caused the growth of industrial and service sectors after losing the majority of oil revenues to South Sudan (African Development Bank, 2019). Furthermore, this situation became even worse for Sudan as tensions with South Sudan over disagreeing on transit fees for oil from South Sudan, along with conflicts and armed violence negatively affected social and economic development in the country (UNESCO, 2018). Besides, Sudan averaged the second-highest inflation rate in Africa at 43.5% behind South Sudan, which averaged a gravely high rate of inflation at 104.1% in 2018 (African Development Bank, 2019; World Bank, 2019b).

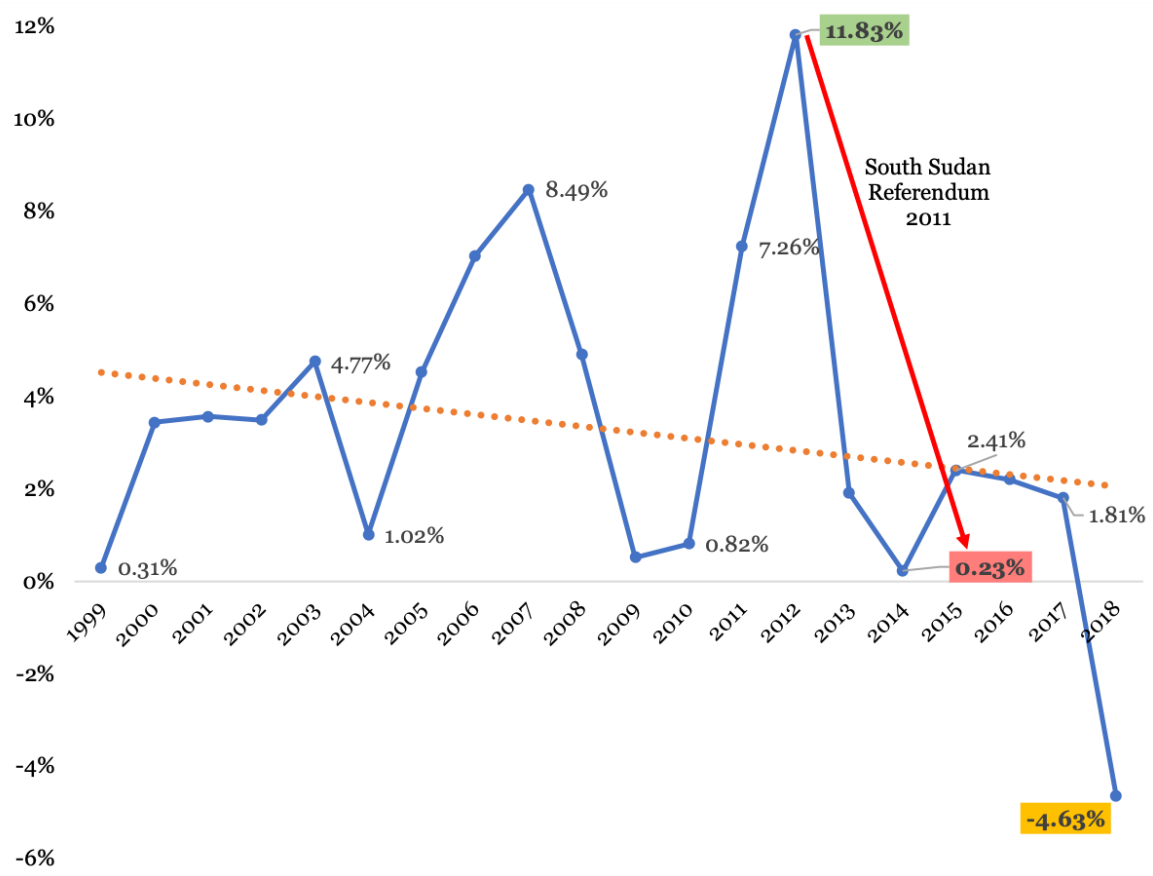


Fig. 2.2 Sudan's GDPPC Growth : 1999-2018
Notes: The Red Line shows the sharp decline in GDPPC growth following the South Sudan secession. The dotted orange line shows the trend in growth over the years.
Data: World Bank (2019b).

2.2.1 Sudan's Performance in Major Global Indexes

The high performance in GDPPC in Sudan between 1996-2001, however, did not translate into a reduction of poverty or improvement in human development (UNESCO, 2018). Sudan's economy was ranked 166 out of 180 countries in the 2019 Economic Freedom index (EFI), and averaged a score lower than both the region and the world (Heritage Foundation, 2019). In the 2019 EFI, Sudan performed low in areas of the rule of law, regulatory inefficiency, and open markets[2]. This could be due to a) Sudan's corruption problems since it is ranked the 172^{nd} out of 180 countries on *Corruption Perceptions Index*; b) informality of the economy due to global sanction with daily transactions taking place with cash, which makes the supervision of money flows and remittances hardened and opens room for illegal activities such as money laundering (Zdanowicz, 2009).

Another important index where Sudan performs scantly is on the 2019 *Fragile States Index (FSI)* (The Fund for Peace, 2019) where it is ranked as the 8^{th} most fragile country –out of 172 countries– in the world. While the EFI by Heritage Foundation (2019) focuses on factors that impact a country's economy, FSI takes a more holistic approach to measure country well-being. FSI is made out of four apparatuses, and a higher score indicates a higher fragility (0-120) (The Fund for Peace, 2019) :

- **Cohesion Indicators:** Security, power of elites, and group grievance (to measure any divide between a particular group and the society).
- **Economic Indicators:** Economic decline and poverty; fairness of development; and human flight/brain drain.
- **Political Indicators:** Legitimacy of elections; enforcement of human rights and the rule of law; and quality of public services.
- **Social Indicators:** Demographic pressures; public health and nutrition; the impact of refugees and internally displaced people.

The components of FSI (The Fund for Peace, 2019) adequately portray how to prosper a country is from various aspects that impact human life. FSI is helpful to visualize a country's standpoint in the world through the lens of national security; economic and institutional quality; and social indicators which all serve to the quality and the functionality of education systems. Figure **2.3** shows the fragility map of countries in Africa using their sui generis FSI scores. As per seen in the map, Sudan

[2]The EFI components are discussed in Appendix **A**

is one of the six countries in Africa that scored 100+; therefore posing as a high alert country. Figure **2.4** shows Sudan's fragility compared to the world and it clearly illustrates the division between the Global North and the South since there are no countries in the Global North that scored over 80 on the FSI; hence none of them are in the 'high warning' to 'very high alert' zone (The Fund for Peace, 2019). Alternatively, those countries that show high levels of stability in the FSI – Finland, Hong Kong, Singapore, Canada, etc.– are concomitantly knowledge capital driven economies (Hanushek and Woessmann, 2015).

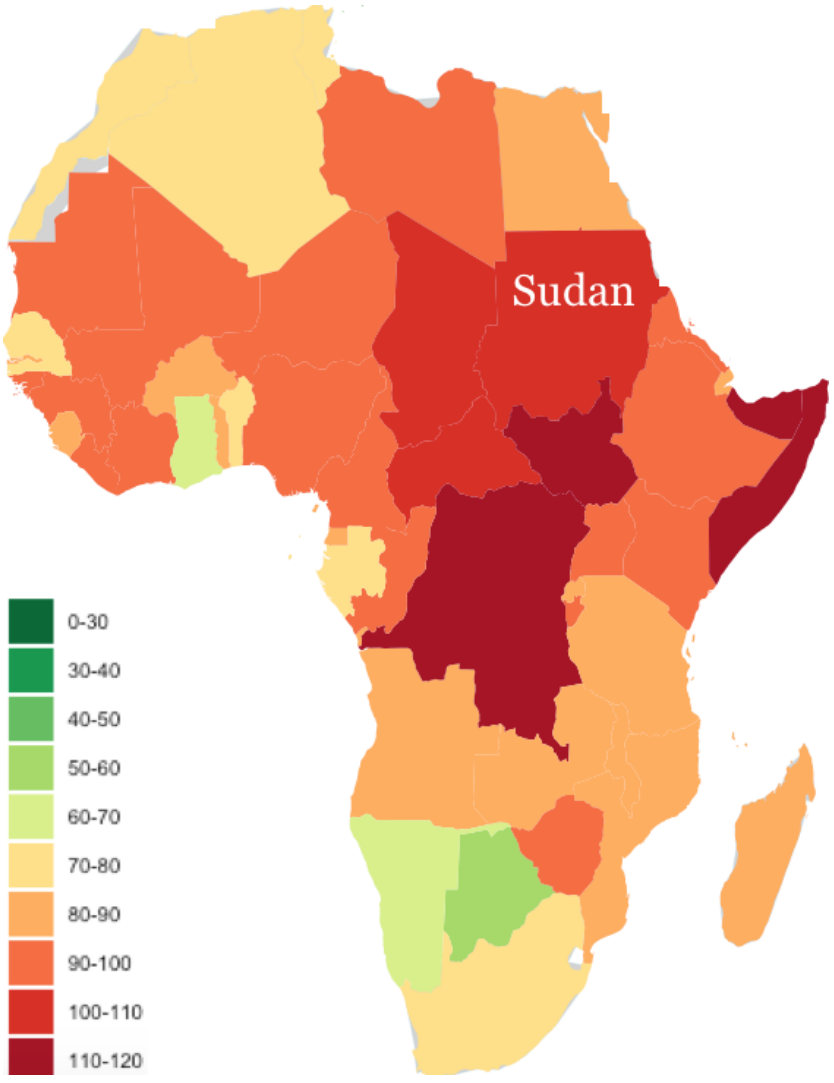


Fig. 2.3 2019 Fragile States Index - Africa. Maximum value= 120. Data: The Fund for Peace (2019) * See Appendix B for plot creation

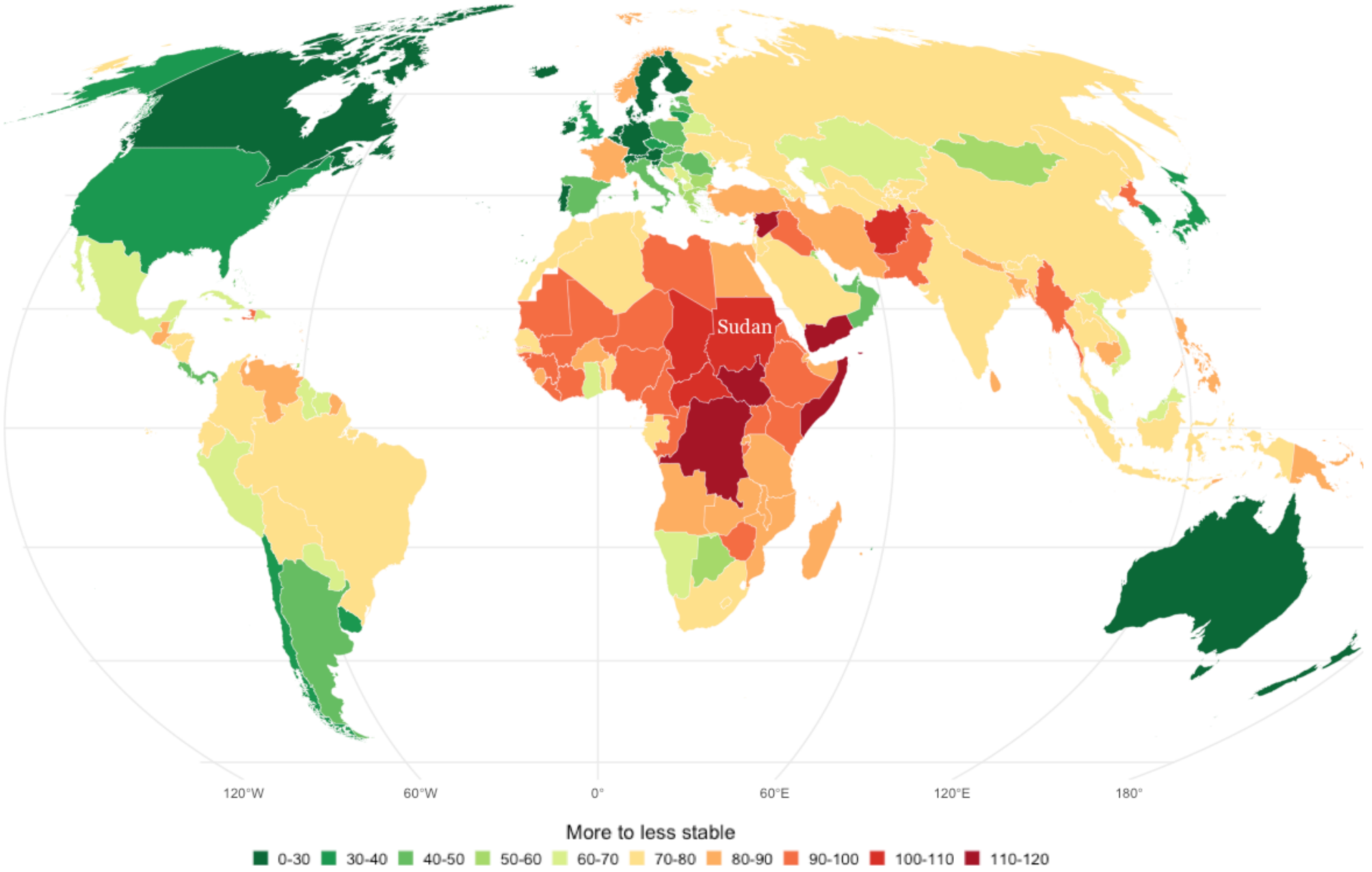


Fig. 2.4 2019 Fragile States Index- World. Maximum value= 120; Data: The Fund for Peace (2019)
* See Appendix B for plot creation

2.2.2 Sudan's Growth Accounting

The final analysis for Sudan is the growth accounting function by Mankiw et al. (1992) to analyze what drove GDPPC growth over the years. Data from Feenstra et al. (2015)'s PWT is useful for such purpose. The scarcity of data is prevalent among underdeveloped countries due to lack of data collection and unwillingness to share data not to reveal exiguous performance. Alas, this problem was persistent in the Penn World Tables 9.1 data set for Sudan as data begins in 1974 in comparison to most countries that have data from 1950. Despite this problem, what obstructed the analysis was the lack of data for '*average annual hours worked by persons engaged*,' which was empty for Sudan for all years. Alternative public data sets such as the OECD (2018); World Bank (2019b), and African Development Bank (2019) also have no data for this variable. As discussed in Section **1.2.1**, the Solow residual, *TFP*, is the residual economic growth unexplained by capital stock, labor force participation, and human capital. Given the average annual hours worked by persons engaged is used as a proxy to measure the labor input contribution, lack of this data also prevents to calculate the TFP portion of the economic growth.

Table 2.1 Economic Growth Accounting for Sudan between 1974-2017 using the Human Capital Augmented Solow Model

	Sudan
Average GDP Growth Rate	5.46
TFP + Labor Input Contribution	2.78
Human Capital Contribution	0.68
Physical Capital Contribution	2.00
Growth from Physical Capital	36.61
Growth from Human Capital	**12.52**

Notes: All numbers are percentages. Growth accounting was done using the derivations in Mankiw et al. (1992).
Data: PWT 9.1 (Feenstra et al., 2015; Zeileis, 2017)

Nevertheless, this analysis was conducted in Table **2.1**. Heavy reliance on growth from physical capital and lack of growth driven by human capital is persistent for Sudan, which is analogous to results for Mexico in **1.2**. Growth from TFP and labor lack explanatory power as the scarcity of data impedes the analysis to find out where the growth came from. Furthermore, given longer working hours do not indicate higher productivity or technological growth, the addition of economic growth from TFP and labor force is not explanatory. Nonetheless, despite experiencing average GDP growth of 5.46%, lack of reliance on human capital again seems to be explaining the plight of Sudan's economy as the results are similar to countries that followed such path.

2.3 The Status of the Sudanese Education System

2.3.1 Historical Context

Historical accounts show that education was mostly informal with an emphasis of Islamic teachings until 1821 when regular schools first appeared under the rule of the Ottoman Empire (Fanack, 2018). Both the increase in the number of public schools and the expansion of privatization in education – which burgeoned in the last three decades – began under the rule of the British-Egyptian Mandate between 1898-1956 (Fanack, 2018). After Sudan's independence in 1956, despite Britain's attempts to convince the government to keep the missionary schools in the South, the "Sudanese nationalists were convinced that missionary education was part of a nefarious divide-and-rule policy" (Johnson, 2017, p 249). The missionaries were expelled out of the country in 1964 following the early establishment of Arabization and Islamization in the country (Johnson, 2017). Johnson (2017) also examines the sizable and rapid expansion of the provision of formal education in Sudan following the independence; however, argues that the returns of education were lower than its expected value given how the "sustained conflict had a devastating effect on education particularly in the south, and [emphasis added] the deviation of funds towards the war to fight the conflict hindered the development of the educational system in the north" [p. 365].

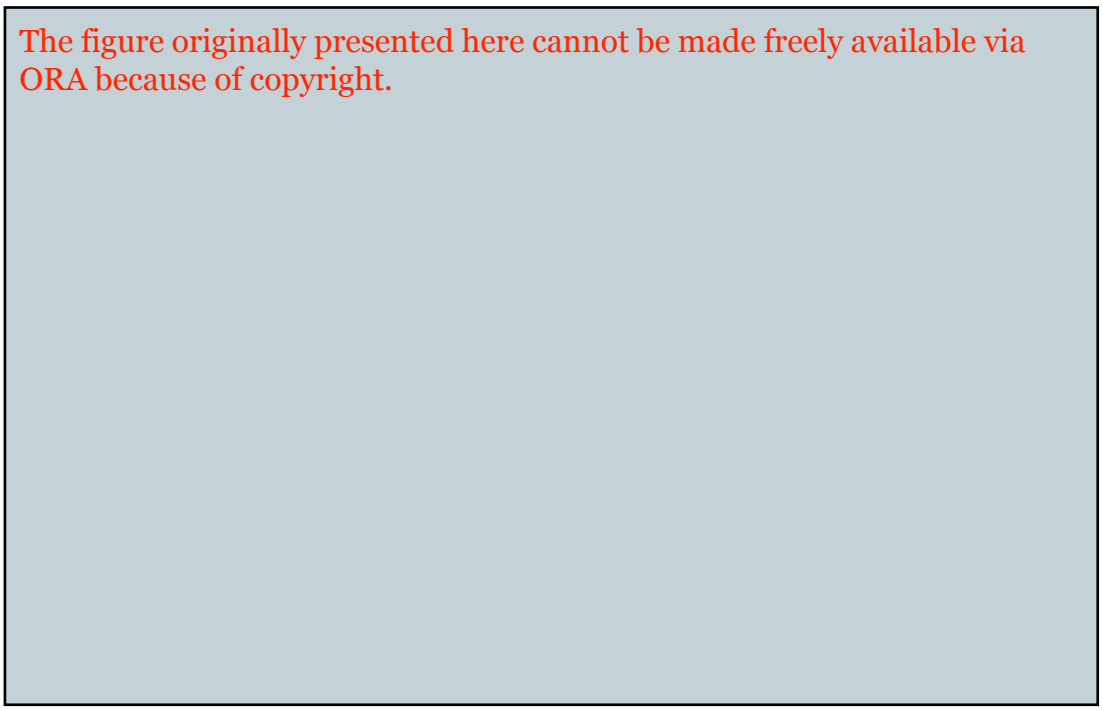


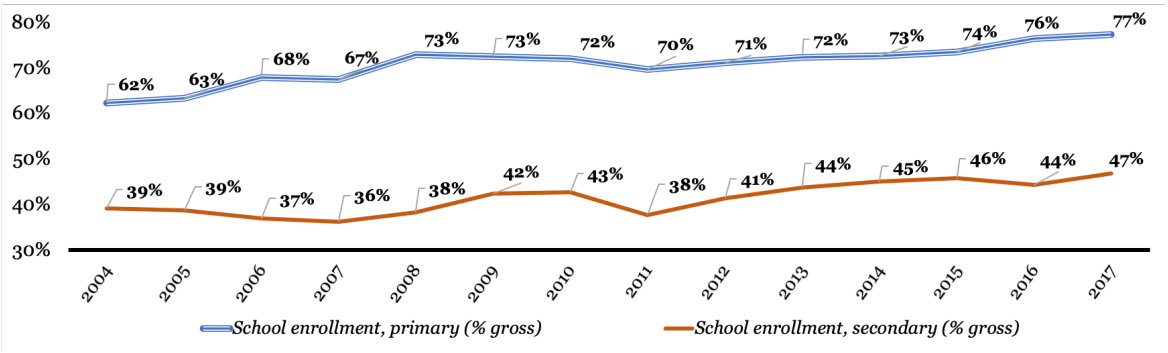
Fig. 2.5 The States of Sudan. Source: African Development Bank (2016)
Note* Highlighted region, Khartoum, is the capital

2.3.2 Gross School Enrollment Trends

World Bank (2012)'s work on the education sector in Sudan still seems to be the most recent and comprehensive study conducted, which provides state-level analyses in the country. These 18 states, which currently make up the current territory of Sudan, are shown in Figure **2.5**. In the study, the period after 2005, when the decades of civil were ended with the Comprehensive Peace Agreement (CPA), is described as conducive for the reconstruction and development of the educational institutions in the cataclysmically affected areas in the country. The evident rise in gross school enrollment rates per state between 2005-2009 is shown in Figure **2.6a**. The positive impact of the CPA on enrollment rates in states that were impacted by the conflict the most such as the North and South Darfur; North and South Kordofan and Western Darfur is seen in Figure **2.6a**. The gains in school enrollment, economic growth (shown in Figure **2.2**, and increase in government spending since last decade; however, were not materialized in socioeconomic factors as poverty is still widespread with close to 50% of the population living below the poverty line (World Bank, 2012).What is more in the socioeconomic setting of Sudan is that around 34% of children in 12 out of 18 states were mentioned underweight in juxtaposition to the Sub-Saharan average (World Bank, 2012).

The figure originally presented here cannot be made freely available via ORA because of copyright.

(a) Change in Enrollment Rates per State Between 2005-2009. Source: Johnson (2017)



(b) School Enrollment Rates in Sudan between 2004–2017. Data: World Bank (2019b)

Fig. 2.6 School Enrollment Rates by Country and State Level in Sudan

Figure **2.6b** rather shows the gross primary and secondary school enrollment rates in Sudan, which seems to follow an upward trend since 2004. The figure shows that the most recent enrollment rate was 77% for primary and 47% for secondary schools in 2017. EQUIP-Sudan (2016); World Bank (2012) and Johnson (2017) argue that one of the main reasons of progress in school enrollment is the Sudanese government to allow international organizations such as the World Bank, UNESCO and the GPE to participate in the decision making, research and funding in education. For instance, the World Bank (2018) maintains a portfolio of around $130 million to manage projects, particularly with a focus on education. One of the major educational projects that the World Bank (2018) carried out in Sudan was the *Basic Education Recovery Project* (World Bank, 2019a), which took place between 2013- February 2019 with a cost of $76.5 million. The World Bank (2018) states the results from this project as:

> The project printed and distributed 22 million textbooks, completed the construction of 1,182 classrooms, and provided grants to 6,399 schools. The total number of student enrolled was 6,020,820, of which the females' percentage was 47% ... Two rounds of **National Learning Assessments** have been developed to assess the quality of education within two grades (three and six) in the Basic Education level [para 2].

The highlighted portion of the above quote is the National Learning Assessment (for 6^{th} grade students) that this thesis uses in the following chapter for data analysis.

2.3.3 The Governance of Sudanese Education System

Another critical factor that has contributed to Sudan's effective educational policy-making – in comparison to the period before 2005 – is the decentralization of power in the governance of the education sector (World Bank, 2012). If implemented correctly and devoid of improper resource distribution, education systems can impressively benefit from decentralization (IIEP, 2018) to address the varying needs of students across the nation. Sudan benefits from decentralized education governance, where the division of decision-making power is majorly dispersed across states and in comparison to more developed countries such as Turkey where more than 94% of the decisions within educational policy-making is taken by the Ministry of Education – which is one of the highest in the world (Gershberg et al., 2012).

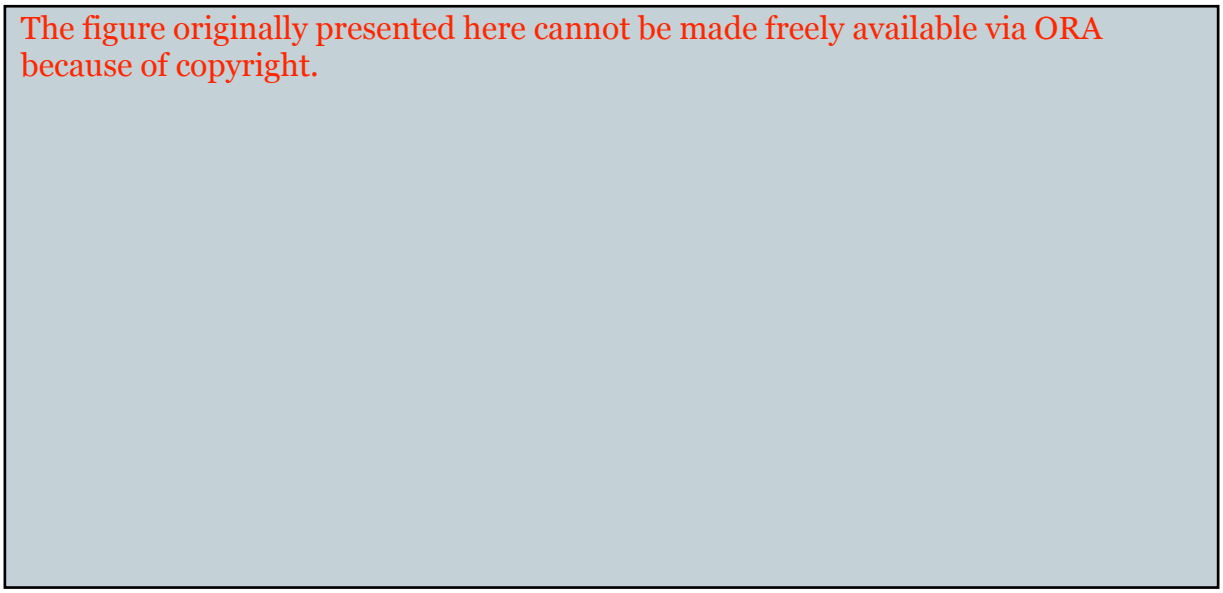


Fig. 2.7 Organization of the Educational Governance in Sudan. Source: UNESCO (2018, p 20)

Figure **2.7** shows the organizational structure of the educational governance in Sudan. UNESCO (2018)'s take on the control that the Federal Ministry of General Education (FMoGE) has over education policy is the following:

> The FMoGE plans, coordinates, and monitors across the three general education levels. Policy development and research, curriculum development, secondary level certification, and teacher qualification also fall under its responsibility. The states and localities are in charge of human resource management, primary level certification, and delivery of pre-school, basic and secondary education [p. 18].

As mentioned in the above quote, the state and local levels of involvement in public education management are quite sizable in Sudan. Figure **2.8** shows the public education spending at the federal and the state level, where the dominance of decision-making over educational spending by states is clearly illustrated. World Bank (2012) writes that following the decentralization of education, educational spending by states showed a significant increase, reaching at 83% in 2009. While the level of decentralization of the education in Sudan might be perceived as a prima facie component of the system, it also creates challenges within the supervision of funds and disparity of the quality of education delivered to students across the nation (World Bank, 2012).

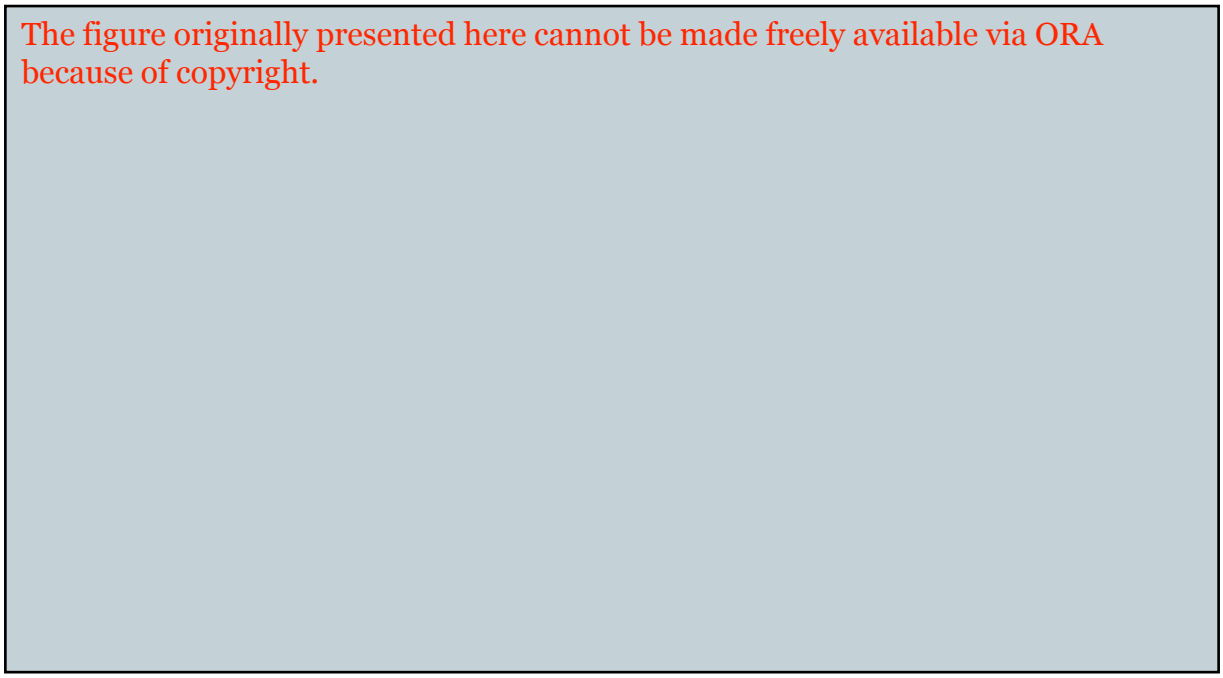


Fig. 2.8 Sudanese Public Education Spending by Institutional Level : 2000–2009. Source: World Bank (2012, p 14)

General public education is 13-years-long, predicated upon three essential levels (UNESCO, 2018), and the current constitution orders the states to deliver free education to all citizens (Fanack, 2018). According to (UNESCO, 2018), the three levels of general public education, which the FMoGE is responsible for are :

1. **Pre-Primary Education**: Children at the age of 4 can start their enrollment in the pre-schools (kindergartens or Islamic schools), which lasts two years. The curriculum is developed at the state level, and pre-primary education is neither compulsory nor free for the citizens.

2. **Primary Education**: Primary level marks the beginning of compulsory education, which lasts for eight years (Grade 1-8) and designated for students at ages 6-13. At the end of Grade 8, students must obtain their *Basic Education Certificate* through an exam. The National Curriculum, which was introduced in 1997, follows five core subjects around religion, mathematics, science, theology, and languages; and, alternatively, aims to provide practical life skills.

3. **Secondary Education**: Students between age 14-16 can attend secondary education and spend three years on "nine core subjects (including Islamic studies, English, history, and mathematics as well as optional subjects in computers, agriculture, commercial science, etc.) [p.19]." The students need to choose between either arts or science as their 3^{rd} year path. The end of secondary education is also the end of compulsory education for Sudanese students, where students sit through the *Sudan School Certificate examination* to both obtain their degree and continue their path to higher education.

2.3.4 Key Challenges of the Education System in Sudan

Despite efforts over the past decade in Sudan to revamp the education system, years of prolonged conflict and poor resource management took its toll on the country; hence, much educational work is ahead (EQUIP-Sudan, 2016; World Bank, 2012). Sudan's literacy rate is currently 75.9% – 83.3% for the male and 68.6% for the female population– (Central Intelligence Agency, 2019), and overall gross enrollment rate is 70.4% (UNESCO, 2018). "However, Sudan still lags behind most of its neighboring countries in terms of universal primary schooling " (UNESCO, 2018, p 73). Furthermore, as per seen in Figure **2.9**, Sudan is the lowest spender on public education as both in the percentage of total public spending (around 13%) and GDP (around 2.7%) in comparison to the analogous countries to Sudan in the region (World Bank, 2012).



Fig. 2.9 Public Education Spending in Africa:2000–09. Source: World Bank (2012, p 163).

One problem where studies show a robust consensus in on the school dropout rates in Sudan (EQUIP-Sudan, 2016; Fanack, 2018; Johnson, 2017; Siddig and Hassan, 2016; UNESCO, 2018; UNICEF, 2016; World Bank, 2012). Keeping students in school is a serious challenge since Sudan averaged dropout rates of 6% of students per grade in primary, and 13% in the first two years of secondary education in 2016 (Siddig and Hassan, 2016). World Bank (2012) finds that "the probability of a student's dropping out before grade 8 is 43% [emphasis added]" [p.4]. Alternatively, as cited in Fanack (2018), Salih (2016) identifies that more than half a million children are at a dropout risk from school, while more than 40% of children at ages between 5-13 have not received any form of education, which roughly equates to three million kids (UNICEF, 2016). The rates of children out of primary school by the state is shown in Figure **2.10**. It seems that those states that were affected by the incessant conflict before the CPA in 2005 such as West and Central Darfur; and West Kordofan are having trouble with enrolling children and keeping them in schools.

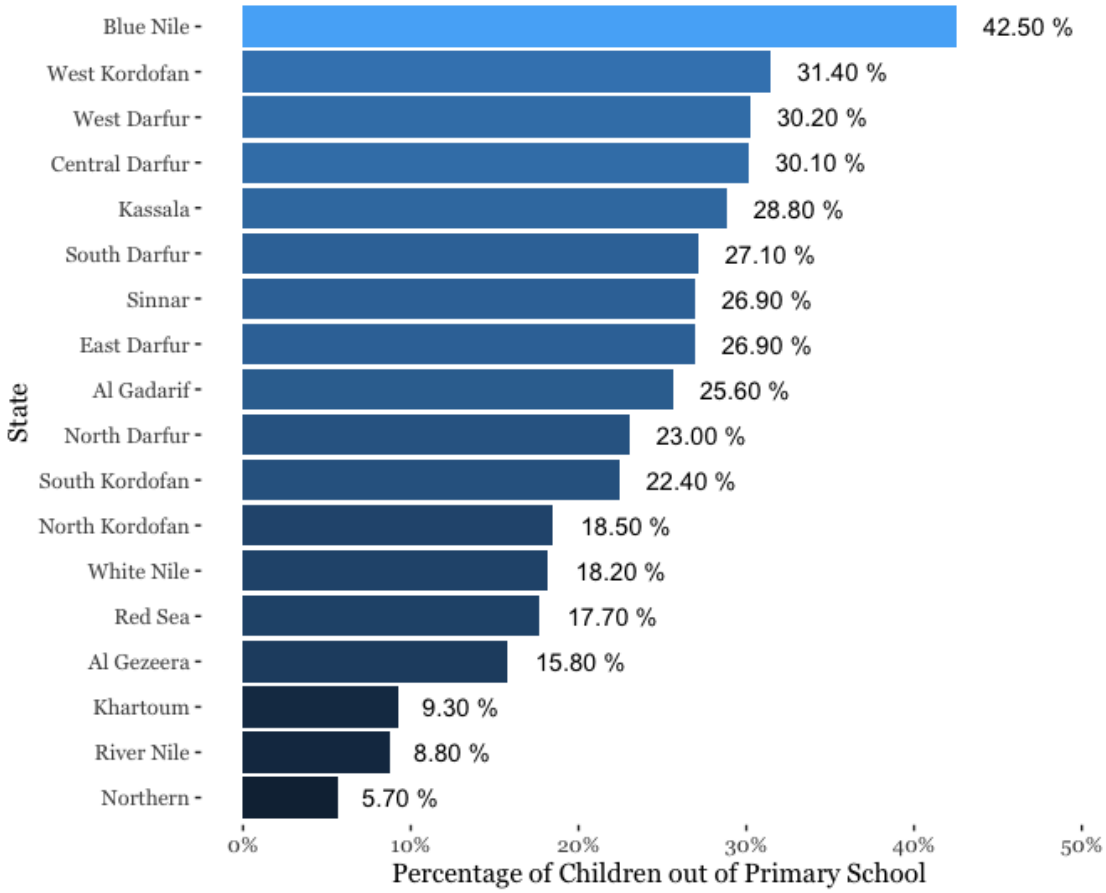


Fig. 2.10 Children out of Primary School in Sudan, by State (2016). Data: UNICEF (2016). * See Appendix B for plot creation

Although the current constitution orders the States Ministries of General Education to provide free education to all citizens, Fanack (2018) argues the reason associated with dropout rates as:

> All universities and almost all primary and secondary schools ask their students to pay registration fees at the beginning of the school year. Sometimes students are also asked to pay other fees throughout the year, which deprives many children of their right to education because of poverty [para 13].

Fanack (2018) also states that a significant portion of families is also unable to afford to serve breakfast for their children, leaving these students no option to either attend to school hungry or instead drop out of school and find a job. Alternatively, World Bank (2012) writes that dropout rates are among the most vital challenges in Sudan

system since many parents do not perceive the amount of investment in education worth more than their children getting a job and providing an additional source of income. Additionally, World Bank (2012) argues that late entry to school by overage children also contributes to their likelihood of dropping out of school as more than 40% of children registered to first grade in 2005 were found to be eight years or older, hence overage. Other factors associated to the high levels of dropout rates are mentioned as "early marriage of girls, child labor, and lack of water and latrines in some schools" (Siddig and Hassan, 2016, p 4). Finally, UNESCO (2018) argues that repetition rates in school serve as a catalyst to Sudan's student retention rates as 15% of students within the first three years of primary education repeated grades.

In summary, some of the important challenges that the Sudanese Education face are :

- Lack of supervision of funds appropriated to states (World Bank, 2012);
- High risk of children dropping out of school (Salih, 2016);
- Burden of the cost families bear to keep their children in school (Fanack, 2018), hence high dropout rates (Siddig and Hassan, 2016; UNICEF, 2016);
- Still-existent problems of lack of water provided to schools, malnourishment, and child labor & marriage (UNESCO, 2018).

Education has still yet to take place in top segments of Sudan's federal budget (World Bank, 2012). It seems that raising student outcomes poses a significant challenge for both the government and the international organizations that endeavor to bring a profound impact on Sudan's education system.

Regarding learning outcomes, Chapter **4** analyzes the most recently conducted learning assessment in Sudan (for 6^{th} grade students) designed by Johnson (2018), which is a part of the Basic Education Recovery Project conducted by the World Bank (2019a). The next chapter (**3**) describes the background of the (Johnson, 2018); the research questions and design; and the methodology used for conducting the analyses in Chapter **4**.

Chapter 3

Methodology

This thesis is designed as an observational study that uses secondary data from Johnson (2018)[1] and obtains purely quantitative methods to answer the research questions. To sum up the questions in Section **3.1.2**, the analyses are designed to test whether the theories from Hanushek and Woessmann (2015) and Cunha and Heckman (2007) hold in Sudan using data from 2018 NLA on 6^{th} grade students on science and mathematics achievement; and survey on students SES background and their access to certain goods & services.

3.1 The Data Set

The Sudanese National Learning Assessment (NLA) is a part of the *Basic Education Recovery Project* (World Bank, 2019a), which took place between 2013- February 2019 with a cost of $76.5 million. As mentioned in Section **2.3.2**, World Bank (2019a) has conducted two NLAs to evaluate educational quality for 3^{rd} and 6^{th} grades.

The first NLA was performed to test reading and mathematics outcomes for 3^{rd} grade students and designed as a "sample-based survey" (Johnson, 2018, p 3). Johnson (2018) writes that between 2014-2015, over 18,000 children in 439 schools that encompasses 18 states in Sudan. The NLA team then conducted the second learning assessment three years later in 2017-2018 on 6^{th} grade students (World Bank, 2019a). The 2018 NLA was also extended to include science and english assessments, along with their "*Early Grade Reading Assessment*" (EGRA) and "*Early Grade Mathematics Assessment*" (EGMA). The questions were designed with the intent to test to what extent students had mastered the education curriculum by the time they reached to

[1]Usage of secondary data from other sources for supplementary analyses also takes place in this thesis and cited in the text and under the relevant figures and tables.

6^{th} grade. Furthermore, a background survey to some students was developed to gain further insight, particularly on factors that shape their life, such as their access to certain goods, family SES, and teacher perception.

The science section of the Grade 6 NLA included domains such as the human body, plants, and environment, evaporation and condensation. The mathematics section alternatively tested students on concepts such as number identification; basic algebra (addition, subtraction, multiplication, and division); shapes and word problems (Johnson, 2018).

3.1.1 2018 NLA Sample

Given Sudan's size and the decentralization policy it follows within the education system, Johnson (2018) writes these factors challenged to identify a genuinely reflective sample of the population. Therefore, the study performed a three-stage cluster survey design.

While a stratified sampling is a standard and more precise method in survey-based research where the target population relevant to the purpose of the study is divided into groups or "stratas," this method is usually employed when almost all necessary information about these units are available to the researcher (Lavrakas, 2008). However, there are some situations, especially in countries like Sudan where data on the population is scant, where the researcher "can only identify pre-determined groups or clusters of sampling units;" therefore conducting cluster survey designs is beneficial to remedy such situations(Lavrakas, 2008, p 2). These clusters in such these survey designs are defined as the"groups of population elements, constitute the sampling unit [emphasis added], instead of a single element of the population" (Ahmed, 2009, p 2). The main goal of cluster sampling, especially the multi-cluster design, is to draw representative units that would take both macro and micro-level aspects of the population into account (Lavrakas, 2008). The most significant trade-off in terms of utilizing cluster sampling is stated to be cost vs. statistical precision (Ahmed, 2009). While cluster sampling is one of the most cost-efficient and feasible in terms of time and implementation, it may, contrarily, not truly capture the diversity of a given community, and result in higher standard errors in comparison to other sampling methods (Ahmed, 2009).

In NLA's design, the primary sampling clusters were the selection of localities in proportion to the size of schools within each state. The secondary sampling unit was the selection of schools, while the final unit consisted of 20 students in each school were selected at random (Johnson, 2018).

In this report, Johnson (2017, p 4) reports the following formula that was used for the sample size calculation :

$$n = z^2 \cdot p(1-p) \cdot (DEFF) \cdot 18(\epsilon p)^2 \cdot (1-r) \tag{3.1}$$

where:

- $n \implies$ required sample size (number of pupils)
- $z \implies$ z value of a 95% confidence interval in a normal distribution (z=1.96)
- $p \implies$ expected prevalence of important indicators (p=0.6)
- $r \implies$ non-response rate (r=3%)
- $DEFF \implies$ the design effect, (DEFF=2)
- $\epsilon \implies$ the relative margin of error at 95% confidence (RME=0.8)
- $18 \implies$ The number of states in Sudan

Johnson (2017) reports that 743 schools were surveyed in the total number of selected students was 15,840. Figure **3.1** shows the number of different types of schools (governmental and nongovernmental; boys, girls, and mixed) from each of the 18 states.

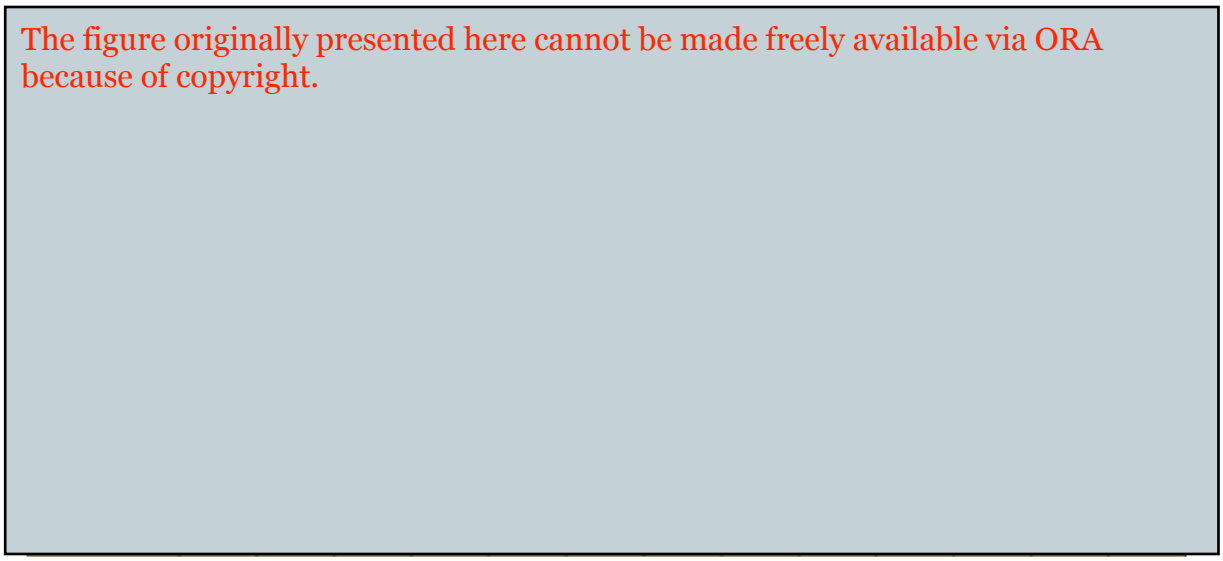


Fig. 3.1 2018 Sudanese NLA's Sample Distribution. Source: Johnson (2018, p 4)

3.1.2 Research Questions

Chapter **1** demonstrates why former human capital augmented economic growth functions using quantity over quality measures failed to capture the impact of education on development. Furthermore, Chapter **1** establishes why the quality of education – hence cognitive skills which can be assessed with achievement in science and mathematics – matters for development; therefore why the factors that could impact cognitive skills should be examined to draw lessons for development and targeted interventions. Therefore, the research questions that this thesis aims to answer are :

1. Why cognitive skills matter for development?
    (a) What is the relationship between achievement in science and mathematics, and poverty rates in the states of Sudan?
2. What explains the variation in cognitive skills?
    (a) What is the impact of family SES and student access to essential goods (electricity, refrigerator, books, computers, etc.) on science achievement?

3.2 Methods and Analyses

3.2.1 Method 1

As per discussed in Chapter **1**, cognitive skills formation is an important predictor of the development of a country. Hence, I analyze the first set of questions in Section **3.1.2** relying on methods from Hanushek and Kimko (2000) and Hanushek and Woessmann (2015) since they are some of the forefronts of the research on cognitive skills and development. Majority of studies discussed in Chapter **1**, such as Hanushek and Woessmann (2015)'s analysis in Table **1.1** and Barro and Sala-i Martin (2004)'s analysis[2] use an educational variable, along with control variables, to predict development within the countries used in their OLS models. Although the analyses for the first set of questions are designed to follow the methods used in these studies, the limitations from the available data prohibit to perform analogous OLS models to predict development in Sudan.

The first limitation for this thesis the inability to adopt methods similar in studies mentioned above since the available data is on a single time point– the year of 2018.

[2]See Appendix **A**

These studies rely on time-series data to build causal models and make causal inferences. Such analyses require a pre and post period to demonstrate a causal pathway for the dependent variable (DV) (Brodersen et al., 2014); henceforward, unable to adopt such a method to analyze development using cognitive skills. One might raise an alternative method of developing OLS models to predict development using poverty rates rather than a causal one; however, the available data also hinders such analysis, which serves as the other limitation. Given that the state-level data only has 18 data points (since there are 18 states), along with a single year point, serves as an unfeasible feature to conduct regression analyses. This is not only a 'micronumerosity' – having a low sample size – (Gujarati and Porter, 2010), but also a central limit theorem (CLT) issue since the CLT conditions that the minimum sample size to explain variance in the DV and possibly having a normally distributed[3] data is 30 data points (Kwak and Kim, 2017). Even if I had access to and included the data from the first NLA conducted in 2014-2015, that would sum up to 36 data points, which would still be considered a low sample size to conduct robust regression analyses (Gelman and Hill, 2018).

Ultimately, I aimed to make the most out of the available data to investigate the linear relationships between science and mathematics achievement and poverty, spending on teachers and good & services by states and relied on the methods from studies such as Hanushek and Woessmann (2015) and Barro and Sala-i Martin (2004). The limitations of data are the reasons why the research question states the word **relationship** rather than **prediction** since the application of any statistical model to predict or show the causal path of an outcome with 18 data points would provide unreliable results. Therefore, I set up bivariate correlation models to analyze the magnitude of these linear relationships by state-level; and hypothesize that:

- **H_0:** There will be no linear relationship between science&math scores and poverty rates across the 18 states
- **H_1:** There will be a linear relationship between science&math scores and poverty rates across the 18 states

I expect the results from these two different correlations to produce similar results to Hanushek and Woessmann (2015)'s findings: there should be a robust negative correlation between science&math scores and poverty rates.

[3]Normality is one of the assumptions of a linear regression (Gelman and Hill, 2018; Gujarati and Porter, 2010)

3.2.2 Method 2

After investigating the first research question, I turn to the NLA data set to investigate the factors that could impact science&math outcomes, as Chapter **1** extensively shows how vital knowledge capital accumulation for countries' development is and why we should investigate it. Given that the data set consists of science & math outcomes, along with a separate file of a background survey, the question again was how to make the best out of the data and perform robust statistical analyses. Along with the availability and the limitations of the data, which is discussed in further sections, it seemed that conducting an analysis of the impact of family SES and student access to essential goods (electricity, refrigerator, books, computers, etc.) on science achievement would best fit the available variables and that the data set included. In this setup, the analysis would benefit both from investigating factors that could predict cognitive skills, along with testing some of the components of Cunha and Heckman (2007)'s theory on skills formation using data from Sudan, which is :

$$\theta_{t+1} = f_t(h_t, \theta_t, I_t) \tag{3.2}$$

where skills at t+1 is a function of θ_t, current skills; h_t, parental and environmental skills; and I_t, parental investment in skills. Since under this framework (in 3.2), skill formation are theorized to be determined by factors these factors, I analyze whether current skills,by taking subsequent achievement (in math) as a proxy; environmental factors such as students availability to certain goods & services and school type; and parental skills and investment into skills such as their education level and work status, hence their SES, would impact science achievement. It is also important to note that Cunha and Heckman (2007)'s theory is one of the most prevalent theories concerning skills formation that encompasses SES factors.

Many studies that inquire about the impact of SES on educational variables with background surveys seem to employ regression analyses. When the DV, the educational variable of the study, is categorical, then studies, for instance, Delprato and Sabates (2014) analyzing late school enrollment usually employ logistic regression models as it would be a more suitable method (Gelman and Hill, 2018). In the case of predicting achievement using SES, the application of OLS models are common (Abuya et al., 2013; Duan et al., 2018; Osullivan et al., 2005; Taylor and Yu, 2009). Such studies are designed where student achievement figures are regressed in OLS models against factors such as parental education, work status or type, previous or subsequent achievement; and control variables such as gender, school type, or other factors researchers deem

appropriate for investigation. These mentioned studies and others in Section **1.4** report varying findings. One reason for that is due to survey data which the effects on certain variables in different samples could point highly dissimilar results since the variance in variables is low (Gelman and Hill, 2018). Therefore, studies seem to get creative with the SES variables by utilizing statistical or machine learning techniques to combine like variables to create continuous variables and indexes.

Taylor and Yu (2009)'s study, where the authors use the PIRLS 2006 achievement and household survey data to determine the impact of SES on student achievement in South Africa, is one of the examples that employs a machine learning technique to create an index of certain goods & services[4]. The variables they use to build their index are also similar to the ones available in the NLA data, and therefore the OLS models that I build to answer the second main research question draws upon its methods from Taylor and Yu (2009)'s study. In their research, the authors state that constructing an index of student assets is an increasingly common approach, and discuss how studies such as Filmer and Pritchett (2001) put forth robust evidence "that asset-based classifications of households correspond closely to classifications by expenditure, and that asset-based indices are in fact better at predicting educational attainment than expenditure information" (Taylor and Yu, 2009, p 14). Creation of such variable would also enable the OLS models in this thesis to account for the 'parental investments in children's skills' component of Cunha and Heckman (2007)'s theory. I refer to this variable as *Student Assets Index* (SAI).

As discussed by Taylor and Yu (2009), the creation of the SAI is not as straightforward as merely adding up the number of items since this approach assumes equal weights are assigned to each asset without taking into consideration how each variable predicts SES and how much of the variation they explain in the data. Since items that are common among households would explain less variation in the data, equal weights approach to build the SAI would be inappropriate Taylor and Yu (2009).

To tackle this problem, I use a machine learning algorithm called *Principal Components Analysis* (PCA), which is also used by Taylor and Yu (2009), to create the SAI using the Phyton programming language. As mentioned in their study, PCA is helpful to assign the appropriate weights to each of the 11 variables that make up the SAI based on their distribution and how much variance each variable explains within the data; hence a higher SD indicates a greater weight of the variable (Filmer and Pritchett, 2001; Jolliffe, 2007). PCA is also useful to eliminate problems such as multicollinearity, which the OLS models assume no multicollinearity in the data, and ensures linearity

[4]I use the term 'goods & services', 'items' and 'assets' interchangeably in this thesis

within the variables it analyzes, therefore contributing to the robustness of the analysis (Gelman and Hill, 2018; Jolliffe, 2007). PCA does this so as it transforms variables that may be highly correlated and non-linear into linear and uncorrelated variables by extracting the linear combinations with the highest level of explanation (Jolliffe, 2007; Taylor and Yu, 2009). Each of these uncorrelated linear combinations that represent various dimensions in the data is called 'principal components,' and the first of those principal components explain the highest variation in data and that variation decreases as more principal components (PC) are created (Taylor and Yu, 2009). Therefore, I use the first PC to construct the SAI using the following formula from Taylor and Yu (2009, p 15), where they both used the first order PC and this formula to build their index:

$$SAI_i = w_1 \cdot \frac{(a_{i1} - \mu_1)}{sd_i} + \ldots + w_{n+1} \cdot \frac{(a_i - \mu_n)}{sd_n} \tag{3.3}$$

where w_1 is the weight associated with the first asset from the first PC; a_i is the value household$_i$ has for the first asset[5]; μ_1 is the mean of the asset 1 for all the household in the data and sd_1 is the standard deviation of the asset 1 (Taylor and Yu, 2009). In the NLA background survey, students were asked whether if they have access to the following assets: electricity, television, refrigerator, phone, motorbike, car, study table, private room, place to study, computer, books (Johnson, 2018). Henceforward, the SAI with these 11 assets was derived from PCA and calculated using the above formula.

Ultimately, by relying on the methods used in studies mentioned in this section, I build a multivariate OLS model using the below theoretical setup:

$$Y = \beta_O + \beta_1 X_1 + \beta_2 X_2 \ldots + \beta_n X_n + \varepsilon \tag{3.4}$$

to test the hypothesis of:

- **H_0:** If $B_1 = B_2 \ldots B_8 = 0$, then dependent variables do not affect science achievement
- **H_1:** If at least $B_1 \neq 0$ or $B_2 \neq 0 \ldots$ or $B_8 \neq 0$, then dependent variables affect science achievement.

with the following OLS equation.

[5] '1' if the household has the asset; and '0' if it does not

$$\begin{aligned}\text{ScienceAchievement}_t = \beta_0 &+ \beta_1 \text{MathAchievement}_t \\ &+ \beta_2 \text{SAI}_t \\ &+ \beta_3 \text{Father's EducationType}_t \\ &+ \beta_4 \text{Mother's EducationType}_t \\ &+ \beta_5 \text{Father's WorkStatus}_t \\ &+ \beta_6 \text{Mother's WorkStatus}_t \\ &+ \beta_7 \text{SchoolType}_t \\ &+ \beta_8 \text{Gender}_t + \epsilon\end{aligned}$$

3.2.3 Rigor

Rigor in Quantitative Methods

It is evident that the epistemological position of this thesis is positivism, which is regarded as the leading philosophy in educational science that employs quantitative methodology (Gasson, 2004). The ontological position of positivism is realism (Marquart, 2017), which is based on the premise that knowledge shall be obtained objectively without the bias of the researcher, predominantly through empirical and mathematical analyses (Scotland, 2012). Henceforward, knowledge solely resides in objects rather than the researcher's interpretation, and they should aim to extract this knowledge through objects (Scotland, 2012).

Marquart (2017) argues that a study's rigor reflects how sound and precise it is at analyzing the research questions. The positivist approach considers the following standards for the rigor in quantitative methods: generalizability (external validity), internal validity and reliability (Gasson, 2004; Muijs, 2011). Generalizability or external validity refers to the ability of a study's sample size to represent the population it was drawn from (Gasson, 2004). The ability of a sample to conclude for broader groups, hence generalizability, is mentioned as a crucial concept that distinguishes quantitative from qualitative methods (Maxwell, 1992; Trainor and Graue, 2014). Internal validity is a crucial concept of the rigor of a quantitative method as it indicates that a statistical relationship in the study should be demonstrated (Gasson, 2004). Finally, reliability is the quality of statistical analysis to be reproduced irrespective of time or the researcher (Gasson, 2004).

This thesis meets the rigor criteria mentioned above as the sample, discussed in Section **3.1.1** shows how it is representative of the population; the findings discussed in the following sections establish a statistical relationship; and finally, these findings using the same data, calculations and methodology could be reproduced irrespective of time or the researcher.

Rigor/Assumptions of Linear Regression Models

OLS models make several assumptions about the variables used in the data set of study (Gray, 2018). Henceforward, it is incumbent upon the researcher to validate that their framework meets the assumption of an OLS model since validating these assumptions might risk the robustness of internal validity (Gelman and Hill, 2018), which is the basis of the rigor of quantitative methods (Gasson, 2004). While running a regression is as simple as writing a single line of code in software such as Phyton or R, investigating the regression model thoroughly and ensuring the existence of the assumptions is far more critical since the findings of regression could be due to various reason other than statistical significance.

These assumptions, also widely known as the *Gauss-Markov Assumptions* (Dodge, 2010b), are (Gelman and Hill, 2018; Gray, 2018):

1. **Linearity**: The linear regression model assumes that there is a linear relationship between the DV and the independent variable(s)

2. **Residuals**: The residuals should form no correlation, and the mean of residuals should approximately be zero.

3. **Normality**: The errors should normally be distributed[6].

4. **Homoscedasticity**: Also known as the homogeneity of variance. The variance of the residuals on the predicted value should be homogeneous, hence the same for all predicted values.

5. **Autocorrelation**: Also known as *Serial Correlation*, this problem occurs when a variable is correlated with its time-lagged version. OLS models assume that the variables are devoid of autocorrelation.

6. **Multicollinearity**: This problem occurs when independent variables are highly correlated among themselves and should not take place in an OLS model.

[6]The normality assumption is often confused with having normally distributed independent variables, which is not the case for the Gauss-Markov Assumptions (Gelman and Hill, 2018)

3.3 Research Ethics

This thesis complies with the ethical guidelines outlined by the Central University Research Ethics Committee (CUREC). As cited in CUREC (2019)'s website: "In general, if the data has been previously collected, and cannot be traced back to identifiable individuals, you need not obtain ethical approval." The data set I use in this thesis is previously collected (between 2017-2018) and thoroughly anonymized along with no information where I can track any of the participants or gain any knowledge about themselves. The data was made available securely by the World Bank and the author, Johnson (2018). The supervisor of this thesis also confirmed that no CUREC approval was needed.

I was provided with separate data files for science and mathematics achievement; and the background survey. I use the student ID numbers that were randomly created merely for the NLA and untraceable to any student, along with 'school type' and 'the state of the school variables' to make sure that almost all variables were retained when I merged the data files using Phyton programming language.

Along with complying with CUREC's guideline, any researcher utilizing statistical and econometric tools to analyze data should endeavor to create the best models. I believe that merely inundating OLS models with loads of variables without deliberately thinking on the ways to improve the model undermines the researcher's responsibility. When working with essential data such as from Johnson (2018), the researcher should feel responsible to provide the most reliable outcomes since these findings could well influence policy which could impact millions of people. In the case of this thesis, more than six months were spent on data analysis from cleaning, manipulation, performing machine learning techniques, and running the countless amount of models to extrapolate the ones that fit the best around the research questions.

Chapter 4

Results

4.1 Why Cognitive Skills Matter for Development?

As per mentioned in Section **3.2.1**, I build correlation plots to investigate the first research question mentioned in Section **3.1.2**. Further state-level achievement and other correlation plots are shown in Appendix **A** and were excluded per word limit purposes. Before analyzing the correlations, I turn to Table **4.1** to point out the successful vs. need-to-improve states.

Table **4.1** shows how the Sudanese states performed in the science and mathematics sections of the 2018 NLA; what percentage of their educational budget they allocate on teachers and goods & services; and hardship figures of the percentage of the population living under poverty and out of school. As per seen in table **4.1**, Al Gezeera, Northern and Khartoum – ranked in that order – were the top three highest-performing states in the science and mathematics outcomes. Conflict-affect states discussed in Section **2.1** such as the Kordofan and the Darfur states, except North Darfur and North Kordofan, were some of the worst performers. The results from the NLA 2018 show the clear divide between the states located in the North vs. South (see Fig **2.5**).

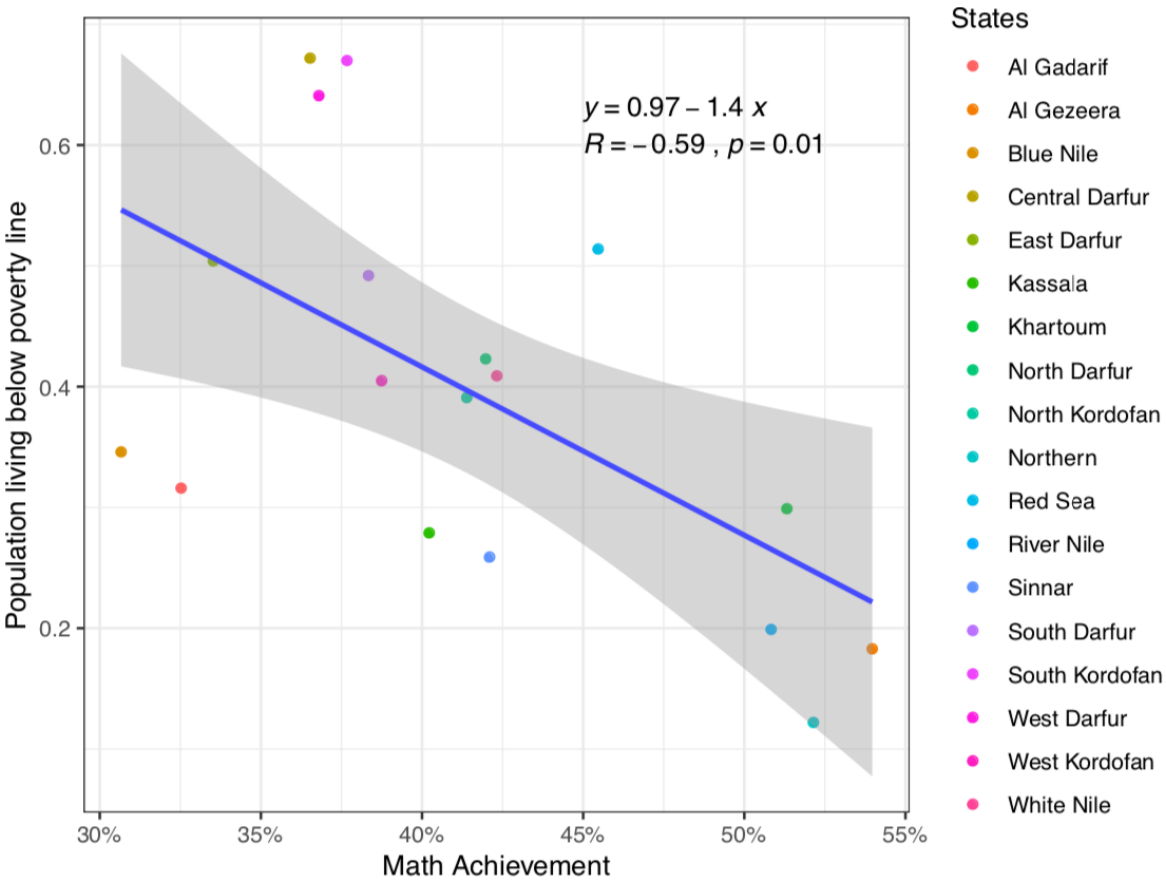
One interesting aspect between the high vs. low performing states is their educational spending figures. For instance, the 2^{nd} highest-ranked state in both math&science outcomes, the Northern state, also spent the highest on goods &services (26% of its educational budget) out of all 18 states. It seems that the states that allocated only around 1% of their education budget for goods & services also performed low in science & mathematics . Given Sudan's problems with corruption as described in Section **2.2.1** (Transparency International, 2018), merely allocating almost all resources to teachers might not yield to the best outcomes in terms of increasing the quality of education. The lack of spending on goods & services creates an additional burden on parents as

World Bank (2012) reports "that households pay a large share of school running costs in addition to other costs, such as uniforms, textbooks, and meals" [p. 7], despite the Sudanese constitution mandates free basic education. Further plots that portray the relationship between educational spending and outcomes are shown in Appendix **A**.

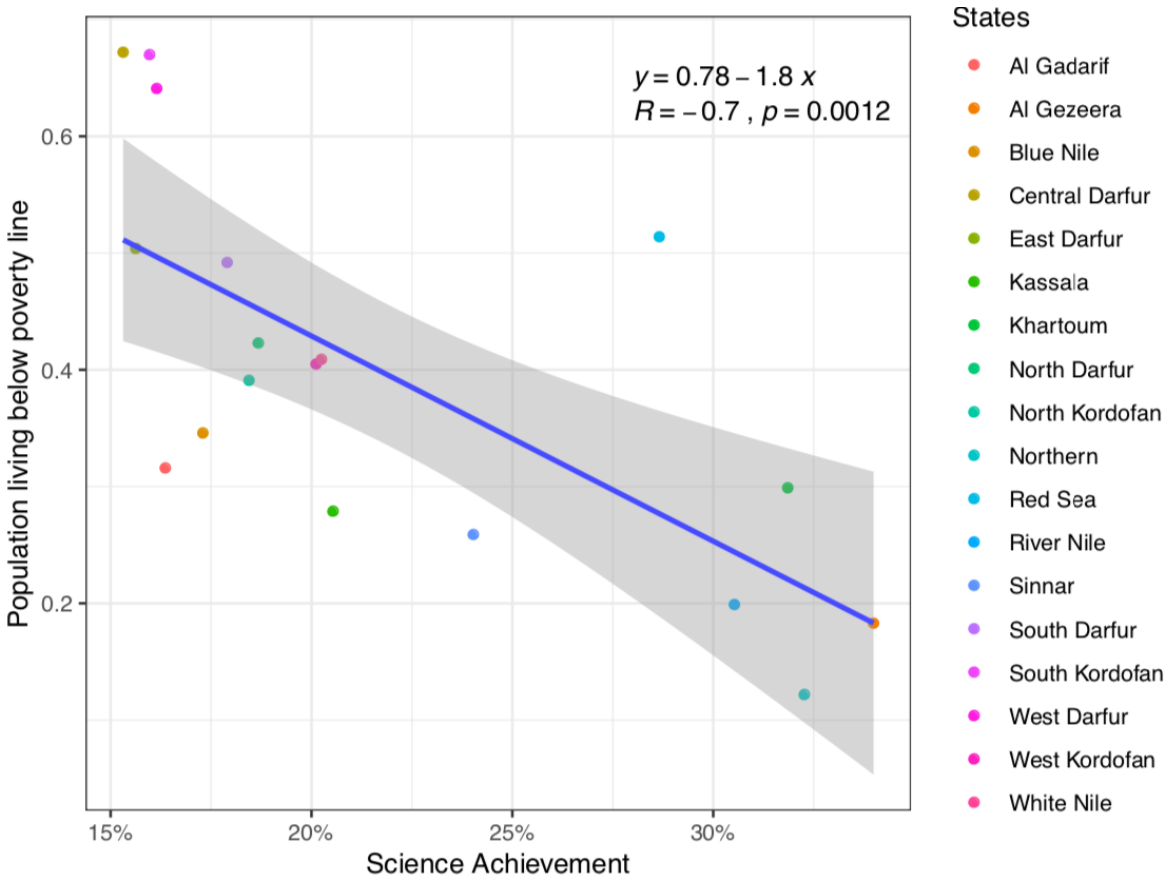
Table 4.1 Achievement and Educational Spending in Sudan, by State

States	Spending		Achievement		Hardship	
	Teachers	G&S	Science	Math	Poverty	Out of School
Al Gadarif	93%	5%	16%	33%	32%	26%
Al Gezeera	92%	6%	34%	54%	18%	16%
Blue Nile	92%	1%	17%	31%	35%	43%
Central Darfur	98%	1%	15%	37%	67%	30%
East Darfur	93%	1%	16%	34%	32%	26%
Kassala	95%	1%	21%	40%	28%	29%
Khartoum	88%	8%	32%	51%	30%	9%
North Darfur	92%	5%	19%	42%	42%	23%
North Kordofan	85%	11%	18%	41%	39%	19%
Northern	72%	26%	32%	52%	12%	6%
Red Sea	90%	2%	29%	45%	51%	18%
River Nile	92%	3%	31%	51%	20%	9%
Sinnar	97%	1%	24%	42%	26%	27%
South Darfur	97%	1%	18%	38%	49%	28%
South Kordofan	98%	1%	16%	38%	67%	22%
West Darfur	90%	2%	16%	37%	64%	30%
West Kordofan	96%	1%	20%	39%	41%	31%
White Nile	89%	8%	20%	42%	41%	18%

Notes 1. Spending figures are percentages from the total budget for education per. 2. G&S stands for goods and services. 3. Poverty column is the population under poverty rates in 2015
Source Achievement:Johnson (2018); Spending & Poverty :Siddig and Hassan (2016); Out of School : UNICEF (2016)



(a) Mathematics Achievement & Poverty



(b) Science Achievement & Poverty

Fig. 4.1 Correlations between Achievement and Poverty in Sudan. Achievement: Johnson (2018); Poverty: Siddig and Hassan (2016)

Figure **4.1** shows the correlations between math&science scores and the population living under the poverty rate. The reason poverty rates were selected as an indicator for development is related to Sudan's significant issues with corruption (Transparency International, 2018) and income distribution. The latest data (in 2009) on income distribution in Sudan shows that highest 20% in Sudan holds 42.4% of the national income, that percentage is only 6.8% for the lowest 20% World Bank (2019b). In an environment where the brethren of elites hold the majority of the national income and not distributed fairly, along with serious corruption issues, GDPPC might not be the most fitting indicator of inclusive growth, which is vital for development (Acemoglu and Robinson, 2012). Therefore, investigation of poverty rates, which are vital development factors, seem to be the more appropriate measures to development in Sudan as states it could well be argued that GDPPC might not reflect well-being in states like South Kordofan where 67% of the population lives under poverty.

Both correlations in Figure **4.1**, with a sample size of 18 (since there are 18 states in Sudan) show robust relationships. The correlation equations; correlation coefficients (R) and p-values are also shown in the figures. Starting with Figure **4.1a**, there is a clear and moderately strong negative relationship between poverty rates and mathematics achievement at R= –0.59, with a significance of p=0.01. The correlation equation, alternatively, estimates that a percentage point increase in mathematics achievement indicated a 1.4 percentage point decrease in the population living below the poverty line; 59% of the variation within these two variables are explained in the correlation. However, it is essential to note that the equation merely reports the change within 18 variables; it does not infer any causation as it is commonly known that correlation does not indicate causation (Gelman and Hill, 2018). The equation shows how two points are in a relationship with one another. Therefore, it can be concluded that mathematics achievement and poverty form a moderately strong negative correlation with an R=59%, and that relationship is significant at p=0.01.

Figure **4.1b** additionally shows the correlation between science achievement and population living below poverty line. That negative relationship, at the same sample size of 18, is even stronger than the one shown in Figure **4.1** at R= –0.7, with a significance of p<0.01. The correlation equation in figure **4.1b** estimates that a percentage point increase in mathematics achievement indicated a 1.8 percentage point decrease in the population living below the poverty line, holding the same standards as mentioned in the above paragraph.

Henceforward, I can reject the null mentioned at the end of Section **3.2.1**; therefore accept the alternative hypothesis that there is a linear relationship between science&

mathematics scores and poverty rates across the 18 states. Furthermore, the expected signs mentioned in Section **3.2.1** also held true in the correlation models.

4.2 What Explains the Variation in Cognitive Skills?

To investigate the hypothesis in Section **3.2.2**, one bi-variate, and four multivariate OLS models were implemented. As mentioned earlier, the DV in all OLS models is the 2018 NLA science achievement from Johnson (2018). The expected initial signs for mathematics achievement; status of parental work; parental education type; and SAI are positive while I hold no expectations or priori assumptions for gender and school type. The sample size, N, used in the models are mentioned in Table **4.2**.

Assumptions of OLS models, as listed in Section **3.2.3** were checked before making statistical inferences. First and foremost, the residuals showed robust normality with a visual examination in the normal probability plot where residuals lie "almost exactly along the diagonal" Field (2018, p 425). The mean of residuals was substantially very small and close zero (-6.785347 e^{-17}). Variance Inflation Factors (VIF) were well below 10 and the tolerance statistics more than 0.2, indicating no problems regarding multicollinearity (Field, 2018; Gelman and Hill, 2018).

The fact that the residuals in the normal probability plot lie almost precisely along the diagonal also reflects that the assumption of homoscedasticity has not been violated (Field, 2018; Gelman and Hill, 2018). Besides, given there is no pattern or linear relation between both the regression residuals and standardized residuals; and predicted value shows that the assumption of homoscedasticity was not validated(Gray, 2018). The relationship of no discernible pattern and non-linearity in the graph of residuals, along with the Durbin-Watson statistic of 1.78 indicate that the model does not face a problem of serial correlation, hence confirming the independence of the residuals (Field, 2018). Furthermore, when the first time period lag (lag 1) of the DV was added as an independent variable to the model, it showed no relationship with the DV, therefore serving as an additional proof for the nonexistence of autocorrelation. Alternatively, the Runs Test for Randomness – an additional check for autocorrelation – produced insignificant p-value, indicating a failure to reject the null hypothesis of the randomness of residuals; therefore confirming that residuals are not autocorrelated. All of the OLS assumption checks are also shown in Appendix **A**. It can be concluded that the model accounts for all the OLS assumptions. Hence the following interpretations based on the results reported in Table **4.2** are statistically valid.

Table 4.2 OLS Models 1-5

	Dependent variable:				
	Science Achievement				
	(1)	(2)	(3)	(4)	(5)
Current Skills					
Math Achievement	0.448***	0.419***	0.415***	0.433***	0.422***
	(0.025)	(0.025)	(0.025)	(0.026)	(0.026)
Environmental					
SAI		1.450***	1.170***	1.157***	0.584*
		(0.265)	(0.279)	(0.297)	(0.346)
School Type					
Nomads			4.093	4.031	3.557
			(6.683)	(6.749)	(6.687)
Private			11.252***	10.829***	9.182**
			(3.810)	(3.977)	(4.001)
Public			7.958**	8.089**	7.616**
			(3.504)	(3.665)	(3.665)
Student Specific					
Gender (Girls)			1.226	0.904	0.967
			(0.963)	(1.003)	(1.007)
Parental Factors					
Do Parents Work?					
Father (Yes)				−1.505	−1.219
				(2.167)	(2.357)
Mother (Yes)				−2.160*	−2.953**
				(1.120)	(1.175)
Father Education Type					
Illiterate					1.854
					(1.649)
Secondary					2.811**
					(1.417)

University					1.157
					(1.808)
Mother Education Type					
Illiterate					−2.098*
					(1.427)
Secondary					2.662*
					(1.569)
University					4.351**
					(1.922)
Constant	4.112***	5.385***	−3.103	−1.692	−2.470
	(1.232)	(1.225)	(3.671)	(4.327)	(4.542)
Observations	582	582	582	534	518
R^2	0.354	0.386	0.398	0.421	0.445
Adjusted R^2	0.353	0.384	0.392	0.412	0.430
Residual Std. Error	11.785	11.503	11.424	11.370	11.201
F Statistic	317.785***	181.709***	63.396***	47.682***	28.744***
Degrees of Freedom	(1; 580)	(2; 579)	(6; 575)	(8; 525)	(14; 503)

Note: Values in parentheses are standard errors. *p<0.1; **p<0.05; ***p<0.01
Data: Johnson (2018)

Regression outputs are in Table **4.2**. I first explain how adding different variables in each step affected the regression model, then analyze the final model in the second next paragraph. This is for two reasons. First, it enables me to analyze Cunha and Heckman (2007)'s theory along with the regression equation shown in Section **3.2.2**. Besides, Model 5 explains the most variation in the DV. Second, when these models were tested for the Akaike's Info Criterion (AIC), which estimates the quality of statistical models in comparison to the other models created using same data (Akaike, 1974), the AIC value was the lowest for Model 5, indicating that it has a better fit concerning the other models (Field, 2018). Building upon this, Model 1 shows that 35.4% of the variation in science scores[1] can be explained by achievement in math, and that relationship is

[1]Also referred to as "DV" throughout the paragraphs

significant across models at p<0.001 (F ((1; 580) = 317.785, R^2 = 0.353[2]). While for Model 1, one percentage point increase in mathematics scores predicts a percentage point increase of 0.448 in science scores, that effect only goes down to 0.422 percentage point increase when all variables are added to the regression output in model 5; the level of significance is at p<0.01 across the models for mathematics scores.

In model 2, the SAI is added to the regression and that is also significant at p<0.001 (F ((2; 579) = 181.709, R^2 = 0.384). In Model 2, a unit increase in the SAI is associated with a 1.45 increase of percentage points in science achievement. When the school type and gender dummy variables and were added, attending to private school is associated with an 11.252 percentage point increase in science achievement significant at p<0.01 (F ((6; 575) = 63.396, R^2 = 0.392); and attending to public school indicates a 7.958, significant at p<0.05 (F (6; 575) = 63.396, R^2 = 0.392). While the gender dummy variable was insignificant, it reports that being a girl student was associated with a 1.226 percentage point increase in science scores. This is a promising result given the widespread egalitarian culture towards women in Sudan, along with harsh enforcement of Sharia Law, which limits women in many aspects of life in the country (Mutiga, 2014). Since the Sudanese education system does not enforce anti-discriminatory laws and regulations, educators can decide on whom they prefer to enroll solely based on their views without facing any consequences (Equal Rights Trust, 2014)

There are interesting results when parental factors added to the regression in Model 4 and 5. When the dummies for whether the student's father and mother works in Model 4, the condition of a student with a working mother predicted a negative impact on science scores, significant at p<0.1. That relationship is more robust as it is significant at a p<0.05 when the control dummies for father's and mother's education type are added in Model 5; and the condition of a student's mother working predicted a 2.953 decrease in science scores (F (14; 503) = 28.744, R^2 = 0.430). Furthermore, while only father's secondary education forms a statistically significant relationship with the DV at p<0.05, and predicts a 2.811 percentage point increase in the DV; all levels of mother's education formed a significant relationship. First and foremost, mother's illiteracy predicts a 2.098 percentage point decrease in the DV (p<0.1, F (14; 503) = 28.744, R^2 = 0.430). On the other hand the effect of mother's education on the science scores increases as secondary and university education dummy variables indicate a 2.662 (p<0.1) and 4.351 (p<0.05) percentage point increase in science scores respectively (F (14; 503) = 28.744, R^2 = 0.430).

[2]Only adjusted R^2 are reported

In conclusion, for the results, five different OLS models used in this paper were able to reject the null and confirm the alternative hypothesis written in Section **3.2.2**. Considering how well models used in this paper fits the data, the first three models reports a moderate relationship with the DV according to Muijs (2011)'s guide[3] while Model 5 shows a strong-moderate relationship with the DV.

[3]A model with an adjusted R^2 value between 0.3- 0.4 has a moderate relationship; 0.4-0.5 has a strong-moderate relationship

Chapter 5

Discussion and Conclusion

5.1 Further Investigation of Mother's Work Type

One of the most unexpected results was the impact of a working mother on science achievement in Table **4.2**. It would not be surprising had one intuitively assumed that a working mother would bring a positive impact on a student's achievement by bringing an additional source of income, or having qualifications to get a job at the first place.

The importance of Chapter **2** therefore emanates following this unexpected result. It is vital to capture various aspects of culture; political and economic systems to make sense of such results as Model 5 in Table **4.2** reports. As mentioned earlier, the widespread egalitarian views on women in Sudan, along with harsh enforcement of Sharia Law, limits women in many aspects of life in the country (Mutiga, 2014). Equal Rights Trust (2014) provides an in-depth analysis of the laws and court cases in Sudan and argues that women face significant discrimination in all segments of society, such as access to education and jobs. Views such as women should be housewives are stated to be prevalent within the society Equal Rights Trust (2014).

Figure **5.1** shows the science achievement of students plotted against the work type of their mothers. As shown in the figure, the second-highest average in science achievement belonged to those that had a non-working mother. While the highest average in science achievement was performed by these that had a mother working as a 'clerk,' it is noteworthy to mention that those students only represented approximately 8% of the sample size when that number for the students with non-working mothers was around 50%. Given the level of discrimination against women in the society along with their limited access to quality jobs (Equal Rights Trust, 2014), the reason behind the lower science achievement of those with a working vs. non-working mother could

well be due to low-quality jobs with meager pay. Since it is not known what qualifies for the choices of 'worker,' 'free-work' and 'retired'; those mothers could have no option but to work at these low-quality jobs to bring an additional source of income and give up the time they would elsewhere allocate for their kids. Therefore, the negative coefficient for mother's work status seems to make sense given the types of jobs they perform according to the data from Johnson (2018).

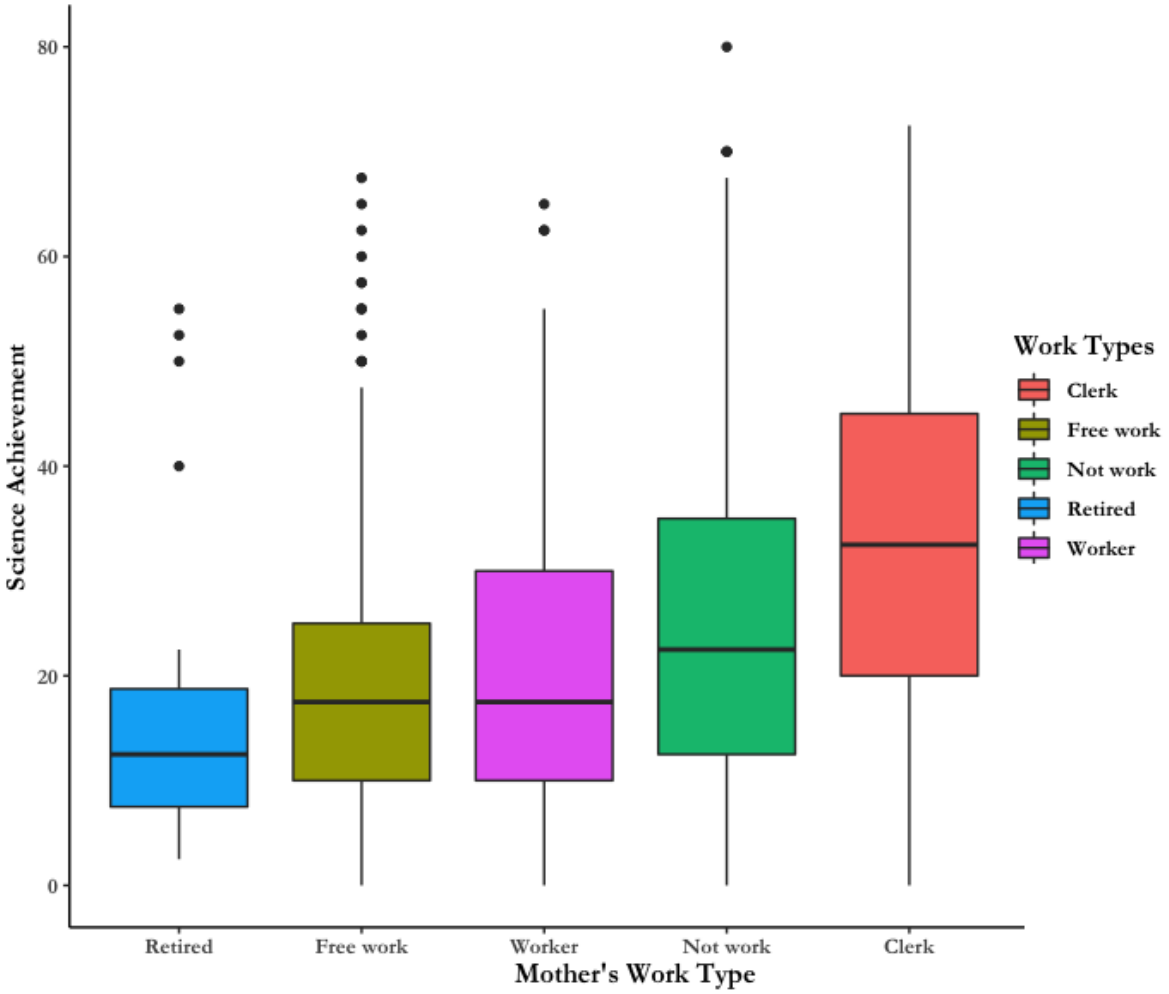


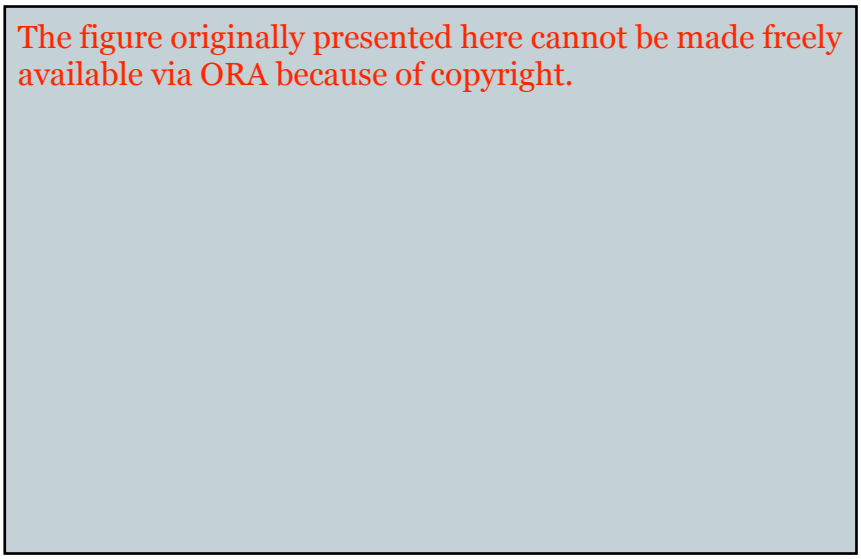
Fig. 5.1 Science Achievement and Mother's Work Type. Data: Johnson (2018)

5.2 Previous Findings vs. the 2018 NLA

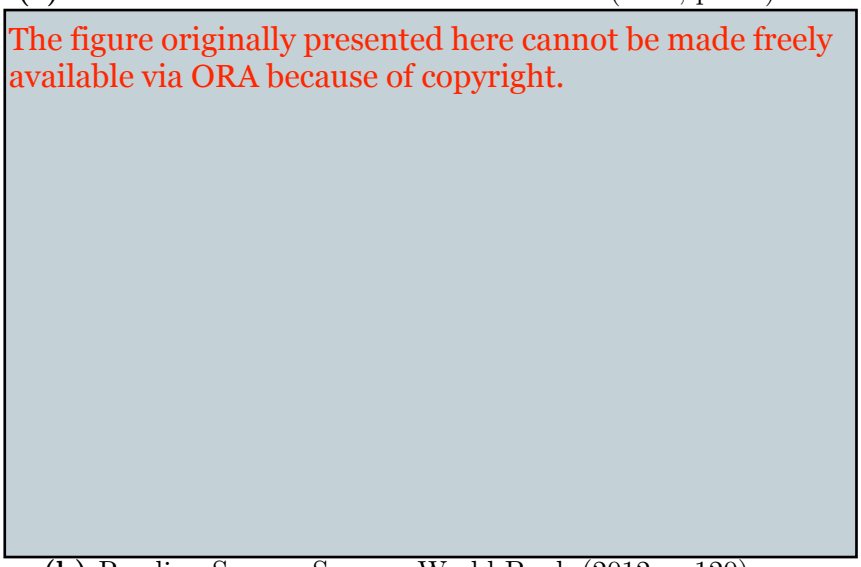
Major studies such as World Bank (2012) and UNESCO (2018) rely on either national-level data made available by the FMoGE. World Bank (2012) also administered a subset of questions from the TIMMS and PIRLS to roughly 4,000 students in Kassala,

North Kordofan, and River Nile in Grade 5, where students on average scored 34% in math and 39% in reading. Regarding the scores, World Bank (2012) states:

> This [results] is low ... With four answer options, if a student simply guessed, he or she would respond correctly 25 percent of the time. The lowest scores in both mathematics and reading were zero, and no student answered all questions on either component correctly. The vast majority of students answered less than half the questions correctly, further showing that [emphasis added] student learning outcomes are very weak [p 119].



(a) Mathematics Scores. Source: World Bank (2012, p 120)



(b) Reading Scores. Source: World Bank (2012, p 120)

Fig. 5.2 2009 World Bank Learning Assessment in Math&Reading in Kassala, North Kordofan, and River Nile (2009). Source: World Bank (2012, p 120)

What makes the NLA unique is that it is the first and latest available data, disaggregated for all the 18 states, on Sudanese learning outcomes (Johnson, 2018). Given how large of a country Sudan is with a decentralized education system, the data can point out how learning varies across states, therefore, enables the analysis to point out successful vs. need-to-improve states, which therefore can open room for further investigation in terms of why some states are performing higher in learning outcomes in juxtaposition to others. Furthermore, one could argue that administering tests from a sample of questions from the TIMMS and PIRLS might not fully capture the learning that takes places in Sudan given the context of the country in terms of its educational and economic development. However, the questions in the NLA were built by the global expertise of the World Bank in conjunction with the FMoGE to analyze learning that takes plan in Sudan appropriately.

5.3 Conclusion and Recommendations

This thesis was designed to answer two questions: a) Why cognitive skills matter for development? b) What explains their variation? Chapter **1** starts a review of the state of the field in exploring the importance of cognitive skills for development; why cognitive skills should be investigated and the methodological progression of the field. To further understand the results from the data analysis in a contextual matter, Chapter **2** provides information on the political economy of education in Sudan. Chapter **3** takes the field to another methodological progression and delineates on the methods used to test the prevalence of claims from Hanushek and Woessmann (2015) on the importance of cognitive skills and Cunha and Heckman (2007) on cognitive skills formation.

The results report essential findings. First, I show that there is a significant negative relationship between poverty and cognitive skills among the states in Sudan in Section **4.1**. As extensively discussed in Chapter **1**, Hanushek and Woessmann (2015) states that cognitive skills matter for development and those countries (in this case states in Sudan) that do not increase their cognitive skills formation fall behind in juxtaposition to the others they do in development. The evidence from these correlation models is relatively in line with Hanushek and Woessmann (2015)'s findings. It can confidently be stated that science & math achievement in Sudan has a negative linear correlation relationship with poverty rates; and that those states that have high achievement in science & math also have lower poverty rates. However, these findings do not claim any causation. While this finding does not hold the same magnitude of statistical power

from Hanushek and Woessmann (2015), it serves as additional support that there is relational support on their claims on the importance of cognitive skills on education from Sudan.

In Section **4.2**, I analyze Cunha and Heckman (2007)'s theory on cognitive skill formation to test how SES impacts science achievement and find that math achievement, SAI, school type (private and public), mother's employment status, father (secondary) and mother education type (illiterate, secondary, university) predicts science achievement with a statistical significance. Explanatory variables such as mother's education usually report different findings in studies. As mentioned in Chapter **1** for instance, Abuya et al. (2013)'s study shows that mother's education only predicted girls' and Osullivan et al. (2005)'s findings only predicted boys' achievement. The OLS models in this thesis elsewhere indicate that each of the three levels of maternal education predicts science achievement in the sample. Furthermore, while the mother's illiteracy predicted a decrease in science achievement, the predictive power of maternal education positively increases as the level of education gets higher. Finally, although the explanatory factors in the regression models did not explain the same level of variation that math achievement did, this is understandable since they are categorical variables. Other studies, such as Duan et al. (2018) that also investigate the impact of SES on achievement, usually find evidence with similar explanatory power.

To summarize, Chapter **1** shows how growth theories and the role of human capital formation in development remain widely-discussed topics in the field of education economics and how previous studies that use school enrollment rates as a proxy for human capital fail to explain the variation in per-capita GDP growth. Furthermore, this chapter also discusses how more recent work in the economics of education such as that of Hanushek and Woessmann (2015)'s study shows that cognitive skills are essential for development; however, how these skills are formed and why should be investigated is less clear. The findings achieved through correlation analyses between state-level poverty rates and science & mathematics achievement show a robust significant relationship. This buttresses the notion with no causal inference that those governments – in this case the states in Sudan – that lack cognitive skill formation also perform lackluster in development. Also in conjunction with Hanushek and Kimko (2000)'s claims, Section **2.3.4** alternatively shows why using school enrollment figures rather than achievement as a proxy for human capital would be implausible given majority of the states have high enrollment levels along with high dropout rates.

Additionally, multivariate OLS models indicate that some components of Cunha and Heckman (2007) theory on skill formation predict science achievement; henceforward

showing to a certain extent what explains the variation in cognitive skills. Therefore, future studies on cognitive skill formation in Sudan would significantly benefit from investigating the parental effect as this thesis shows that they have an impact on learning. For instance, a famous study by Angrist et al. (2001) of a randomized experiment in Colombia where financially disadvantaged families received educational vouchers for their children found many benefits in the post period of the study such as higher achievement and less likelihood of grade repetition. A similar study in Sudan where the impact of giving educational grants to working mothers are for their kids and encouraging or supporting with additional training to apply for higher quality jobs impacts the cognitive skills of their children.

5.4 Limitations

The limitations of a thesis from the resources, time, and word-limit available are not vital to discuss given the awareness of such limitations by the readers; however, there were various limitations of the data available.

Before all else, the NLA data is too nascent with a set of a single time point, therefore limiting the power of findings from causality to prediction: and that prediction is for achievement in a single year. Despite the analysis section of this thesis employed creative methods such as utilizing a machine learning technique of PCA, since that the NLA is not an experiment, one cannot possibly control everything that could impact student achievement. One of the reasons why studies such as Hanushek and Woessmann (2015) and Barro and Sala-i Martin (2004) are considered as groundbreaking studies is due to their high-levels of statistical evidence, which is mainly drawn from time-series models. Therefore, it ought to be acknowledged that various factors could impact science achievement and the final model (5) of this thesis was able to explain roughly 45% of it from 2018 data points.

The lack of time-series data also hinders the explanatory power of the first research question where the relationship between cognitive skills and poverty was investigated. The thesis would have benefited from an analysis similar to the one in Table **1.1** on the states in Sudan. With such limitation, I am only able to show the relationship between cognitive skills and poverty, which one could argue the impact of other factors.

The original purpose of the NLA is to track learning progression, and the data is used for secondary purposes, which is to investigate the research questions of this thesis. Therefore it is understandable that much attention was not paid to the background survey from the people in charge of data collection as a significant portion of the

students did not take the survey. Although the assessments were administered to around 16 thousand people; only 2187 people who took the science test also completed the survey. That number dropped down to 518 when the DVs were added. It is a disappointing factor for a researcher to see a sample size of 16 thousand dropping down to 518 due to data collection mismanagement; however, once again, this thesis is a quantitative observational study using secondary data; therefore, it is typical for such research to not work with data that comprehensively satisfy expectations.

Finally, the lack of available secondary data from leading sources such as the World Bank Indicators and the PWT was another limitation on the analyses. Even analyses as simple as growth accounting was not entirely performed, given the lack of data. The scarcity of data is prevalent among such countries due to lack of quality data collection and government unwillingness to share data not to reveal exiguous performance.

References

Abuya, B. A., Oketch, M., Mutisya, M., Ngware, M., and Ciera, J. (2013). Influence of mothers education on childrens maths achievement in kenya. *Education 3-13*, 41(6):633–646.

Acemoglu, D. (2014). Economic growth: The solow growth model and the data. *Massachusetts Institute of Technology*, page 1–52.

Acemoglu, D. and Robinson, J. A. (2012). *Why nations fail: The origins of power, prosperity, and poverty*. Crown Books.

Acs, Z. J., Estrin, S., Mickiewicz, T., and Szerb, L. (2014). The Continued Search for the Solow Residual: The Role of National Entrepreneurial Ecosystem. IZA Discussion Papers 8652, Institute of Labor Economics (IZA).

African Development Bank (2016). Sudan data portal. http://sudan.opendataforafrica.org/.

African Development Bank (2019). *African economic outlook*. African Development Bank Group.

Ahmed, S. (2009). Methods in sample surveys : Cluster sampling. *Department of Biostatistics: John Hopkins University*, page 1–14.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Al Jazeera (2019). Sudan's ruling generals warn opposition ahead of millions march. https://www.aljazeera.com/news/africa/2019/06/sudan-ruling-generals-warn-opposition-millions-march-190629140621194.html.

Angrist, J., Bettinger, E., Bloom, E., King, E., and Kremer, M. (2001). Vouchers for private schooling in colombia: Evidence from a randomized natural experiment. *The American Economic Review*, page 1535–1558.

Arel-Bundock, V., Enevoldsen, N., and Yetman, C. (2018). countrycode: An r package to convert country names and country codes. *Journal of Open Source Software*, 3(28):848.

Arnold, J. B. (2019). *ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'*. R package version 4.2.0.

Auguie, B. (2017). *gridExtra: Miscellaneous Functions for "Grid" Graphics*. R package version 2.3.

Barro, R. (1991). Economic growth in a cross section of countries. *The Quarterly Journal of Economics*, 106(2):407–443.

Barro, R. (2001). Education and economic growth. *In: Helliwell JF The Contribution of Human and Social Capital to Sustained Economic Growth and Well-Being*.

Barro, R. and Lee, J.-W. (2013). A new data set of educational attainment in the world, 1950-2010. *Journal of Development Economics,*, 104:184–198.

Barro, R. J., Chu, A. C., and Cozzi, G. (2017). *Intermediate Macroeconomics*. Cengage Learning EMEA.

Barro, R. J. and Sala-i Martin, X. (2004). *Economic growth*. The MIT Press.

BBC News (2011). South sudan referendum: 99% vote for independence. https://www.bbc.co.uk/news/world-africa-12317927.

BBC News (2019a). Sudan country profile. https://www.bbc.co.uk/news/world-africa-14094995.

BBC News (2019b). Sudan crisis: Military and opposition agree power-sharing deal. https://www.bbc.co.uk/news/world-africa-48878009, journal=BBC News.

Becker, G. (2002). Human capital. *The Library of Economics and Liberty*. url-https://www.econlib.org/library/Enc/HumanCapital.html.

Becker, G. S. (1962). *Investment in Human Capital: A Theoretical Analysis*, pages 9–49. Volume LXX of National Bureau Committee for Economic Research (1962).

Becker, G. S. (1993). *Human capital: a theoretical and empirical analysis, with special reference to education*. University of Chicago Press.

Bivand, R., Nowosad, J., and Lovelace, R. (2019). *spData: Datasets for Spatial Analysis*. R package version 0.3.0.

Boo, F. L. (2014). *Socio-economic status and early childhood cognitive skills: is Latin America different?* Young Lives.

Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., and Scott, S. L. (2014). Inferring causal impact using Bayesian structural time-series models. *Annals of Applied Statistics*, 9:247–274.

Burke, J. (2019). African union suspends sudan over violence against protesters. https://www.theguardian.com/world/2019/jun/06/sudan-african-union-suspension-military-rulers.

Canarella, G. and Pollard, S. K. (2011). The augmented solow model and the oecd sample. *International Business Economics Research Journal (IBER)*, 2(7).

Cassen, R., Feinstein, L., and Graham, P. (2009). Educational outcomes: Adversity and resilience. *Social Policy and Society*, 8(1):73–85.

Central Intelligence Agency (2019). The world factbook. https://www.cia.gov/library/publications/the-world-factbook/.

Chang, W. (2012). *extrafontdb: Package for holding the database for the extrafont package*. R package version 1.0.

Chang, W. (2014). *extrafont: Tools for using fonts*. R package version 0.17.

Cheng, J., Karambelkar, B., and Xie, Y. (2018). *leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library*. R package version 2.0.2.

Choi, A. S. (2015). What the best education systems are doing right.

Clinton, K. (2004). A useful production function: Cobb-douglas.

Coalition for the International Criminal Court (2009). Omar al-bashir. /urlhttp://www.coalitionfortheicc.org/cases/omar-albashir.

Cobb, C. W. and Douglas, P. H. (1928). A theory of production. In *Proceedings of the Fortieth Annual Meeting of the American Economic Association*, volume 139, page 165.

Cohn, B. B. (2017). Are we human beings or human capital? *BIZCATALYST 360°*. https://www.bizcatalyst360.com/are-we-human-beings-or-human-capital/.

Cunha, F. and Heckman, J. (2007). The technology of skill formation. *American Economic Review*, 97(2):31–47.

Cunha, F. and Heckman, J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43(4):738–782.

CUREC (2019). Research ethics (including curec) :faqs. https://researchsupport.admin.ox.ac.uk/governance/ethics/faqs-glossary/faqs#collapse410591.

DeLong, J. B. (2000). The shape of twentieth century economic history. Working Paper 7569, National Bureau of Economic Research.

Delprato, M. and Sabates, R. (2014). A multilevel analysis of late entry in nigeria. *International Journal of Research & Method in Education*, 38(1):83–99.

Deutsche Welle (2019). Sudan's military ousts president omar al-bashir following protests: Dw: 11.04.2019. https://www.dw.com/en/sudans-military-ousts-president-omar-al-bashir-following-protests/a-48282243.

Dodge, Y. (2010a). *The Concise Encyclopedia of Statistics*. Springer New York, New York, NY.

Dodge, Y. (2010b). Gauss–Markov Theorem. In Dodge (2010a), pages 217–218.

Douglas, P. H. (1976). The Cobb-Douglas Production Function Once Again: Its History, Its Testing, and Some New Empirical Values. *Journal of Political Economy*, 84(5):903–915.

Duan, W., Guan, Y., and Bu, H. (2018). The effect of parental involvement and socioeconomic status on junior school students' academic achievement and school behavior in china. *Frontiers in Psychology*, 9.

Equal Rights Trust (2014). *In search of confluence: addressing discrimination and inequality in Sudan*. Equal Rights Trust in partnership with Sudanese Organisation for Research and Development.

EQUIP-Sudan (2016). Education quality improvement programme in sudan (equip-sudan). https://ec.europa.eu/trustfundforafrica/region/horn-africa/sudan/education-quality-improvement-programme-sudan-equip-sudan_en.

Fanack (2018). Education in sudan: A long history but deeply troubled reality. https://fanack.com/education/education-in-sudan/.

Feenstra, R., Inklaar, R., and Timmer, M. (2015). The next generation of the penn world table. *American Economic Review*, 105(10):3150–3182.

Felipe, J. and Adams, F. G. (2005). "a theory of production" the estimation of the cobb-douglas function: A retrospective view. *Eastern Economic Journal*, 31(3):427–445.

Field, A. P. (2018). *Discovering statistics using IBM SPSS statistics*. SAGE Publications.

Figlio, D. N., Freese, J., Karbownik, K., and Roth, J. (2017). Socioeconomic status and genetic influences on cognitive development. *Proceedings of the National Academy of Sciences*, 114(51):13441–13446.

Filmer, D. and Pritchett, L. H. (2001). Estimating wealth effects without expenditure data-or tears: An application to educational enrollments in states of india. *Demography*, 38(1):115–132.

Fincham, K. (2018). Gender and primary school dropout in sudan: Girls' education and retention in red sea state. *Prospects*, pages 1–16.

Garcia, E. (2014). The need to address noncognitive skills in the education policy agenda. *Economic Policy Institute*.

Gasson, S. (2004). Rigor in grounded theory research: An interpretive perspective on generating theory from qualitative field studies. In *The handbook of information systems research*, pages 79–102. IGI Global.

Gelman, A. and Hill, J. (2018). *Data analysis using regression and multilevel/hierarchical models*. Cambridge Univ. Press, 19 edition.

Georgy, M. (2019). More protesters flood sudan's sit-in to demand civilian rule. https://uk.reuters.com/article/uk-sudan-politics/more-protesters-flood-sudans-sit-in-to-demand-civilian-rule-idUKKCN1RV0UG.

Gershberg, A. I., González, P. A., and Meade, B. (2012). Understanding and improving accountability in education: A conceptual framework and guideposts from three decentralization reform experiences in latin america. *World Development*, 40(5):1024–1041.

Gneezy, U., List, J. A., Livingston, J. A., Sadoff, S., Qin, X., and Xu, Y. (2017). Measuring success in education: The role of effort on the test itself. Working Paper 24004, National Bureau of Economic Research.

Goldin, C. (2014). Human capital. *National Bureau of Economic Research*, page 1–42.

Gray, D. E. (2018). *Doing research in the real world*. SAGE.

Gruber, N. (2017). Why israel does poorly in the pisa exams–perceptions versus reality. *Shoresh Institution for Socioeconomic Research*.

Gujarati, D. N. and Porter, D. C. (2010). *Essentials of Econometrics*. McGraw-Hill/Irwin.

Gunkel, C. (2011). Deutsche sprachpreise: Ein jahr, ein (un-)wort! - spiegel online - einestages.

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

Hanushek, E. A., Piopiunik, M., and Wiederhold, S. (2018). The value of smarter teachers: International evidence on teacher cognitive skills and student performance. Technical report, National Bureau of Economic Research.

Hanushek, E. A., Welch, F., Machin, S., and Woessmann, L. (2011). *Handbook of the economics of education*. North-Holland.

Hanushek, E. A. and Woessmann, L. (2012). Schooling, educational achievement, and the latin american growth puzzle. *Journal of Development Economics*, 99(2):497–512.

Hanushek, E. A. and Woessmann, L. (2015). *The knowledge capital of nations: education and the economics of growth*. The MIT Press.

Hausmann, R. and Rodríguez, F. R. (2014). *Venezuela Before Chávez: Anatomy of an economic collapse*. Penn State Press.

Healy, K. J. (2019). *Data visualization: a practical introduction*. Princeton University Press.

Heckman, J. J. and Mosso, S. (2014). The economics of human development and social mobility. Working Paper 19925, National Bureau of Economic Research.

Heritage Foundation (2019). *2019 Index of Economic Freedom*. The Heritage Foundation.

Hertz, T., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N., and Verashchagina, A. (2008). The inheritance of educational inequality: International comparisons and fifty-year trends. *The B.E. Journal of Economic Analysis & Policy*, 7(2).

Hester, J., Csárdi, G., Wickham, H., Chang, W., Morgan, M., and Tenenbaum, D. (2019). *remotes: R Package Installation from Remote Repositories, Including 'GitHub'*. R package version 2.1.0.

Hijmans, R. J. (2019). *raster: Geographic Data Analysis and Modeling*. R package version 2.9-5.

Hlavac, M. (2018). *stargazer: Well-Formatted Regression and Summary Statistics Tables*. Central European Labour Studies Institute (CELSI), Bratislava, Slovakia. R package version 5.2.2.

IDEAS REPEC (2019). Top 10 percent institutions and economists in the field of education, as of may 2019. https://ideas.repec.org/top/top.edu.html#authors.

IIEP (2018). Decentralization of education management. https://learningportal.iiep.unesco.org/en/issue-briefs/improve-learning/education-system-management/decentralization-of-education-management.

Jeworutzki, S. (2018). *cartogram: Create Cartograms with R*. R package version 0.1.1.

Johnes, G. (1993). *The economics of education*. Macmillan.

Johnson, D. (2017). Sudan: An overview | education in the arab world. *Bloomsbury Collections*, page 349–368.

Johnson, D. (2018). Sudan national leaning assessment 2018: Are more children learning, and are they learning more? *The World Bank Group*.

Jolliffe, I. T. (2007). *Principal component analysis*. Springer.

Jones, C. (2005). The shape of production function and the direction of technical change. *The Quarterly Journal of Economics*, page 517–549.

Kassambara, A. (2018). *ggpubr: 'ggplot2' Based Publication Ready Plots*. R package version 0.2.

Kwak, S. G. and Kim, J. H. (2017). Central limit theorem: the cornerstone of modern statistics. *Korean Journal of Anesthesiology*, 70(2):144–156.

Lavrakas, P. (2008). Encyclopedia of Survey Research Methods.

Liucvaitiene, A., Peleckis, K., Slavinskaite, N., and Limba, T. (2013). Competitive advantage in global market: Theoretical assumptions and assessment options. *Strategic management quarterly*, 1(1):1–10.

Lovelace, R., Nowosad, J., and Muenchow, J. (2019). Chapter 8: Making maps with r.

Luginbühl, F. (2018). Reproducing the economist most popular map of 2017. https://felixluginbuhl.com/mapping/.

Mankiw, N. G., Romer, D., and Weil, D. N. (1992). A Contribution to the Empirics of Economic Growth. *The Quarterly Journal of Economics*, 107(2):407–437.

Marginson, S. (2017). Limitations of human capital theory. *Studies in Higher Education*, 44(2):287–301.

Marquart, F. (2017). Methodological rigor in quantitative research. *The International Encyclopedia of Communication Research Methods*, pages 1–9.

Maxwell, J. (1992). Understanding and validity in qualitative research. *Harvard Educational Review*, 62(3):279–301.

McKinney, W. (2010). Data structures for statistical computing in python. In van der Walt, S. and Millman, J., editors, *Proceedings of the 9th Python in Science Conference*, pages 51 – 56.

Muijs, D. (2011). *Doing quantitative research in education with SPSS*. Sage Publications.

Mutiga, M. (2014). Sudan's 'morality' laws used to punish women, report finds. https://www.theguardian.com/world/2014/oct/02/sudan-morality-laws-women-report.

Nafiu, L. A. (2012). Comparison of one-stage, two-stage, and three-stage estimators using finite population. *The Pacific Journal of Science and Technology*, 13(2):166–171.

National Bureau Committee for Economic Research, U. (1962). *Investment in Human Beings*. The Journal of Political Economy Vol. LXX, No. 5, Part 2 (University of Chicago Press).

National Institutes of Health (2018). Brain basics: Know your brain.

Neuwirth, E. (2014). *RColorBrewer: ColorBrewer Palettes*. R package version 1.1-2.

OECD (2009). The value of people. *OECD Insights: Human Capital*, page 22–37.

OECD (2016). Overview: Policies and practices for successful schools. *PISA 2015 Results (Volume II) PISA*, page 33–46.

OECD (2018). Oecd economic outlook, volume 2018 issue 2. *OECD Publishing*.

Osullivan, V., Walker, I., Chevalier, A., and Harmon, H. (2005). The impact of parental income and education on the schooling of their children. *Working Paper Series*, page 1–30.

Parker, J. (2015). Human capital in the solow model. *REED College*, page 57–65.

Partington, R. (2019). Longer working hours do not mean higher profits, say economists. https://www.theguardian.com/money/2019/apr/15/longer-working-hours-do-not-mean-higher-profits-say-economists.

Patta, D. (2019). Sudan violence prompts trump admin. to send senior envoy to meet military leaders. https://cbsn.ws/2R7QbjL.

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1):439–446.

Pritchett, L. (2009). Does schooling help explain any of the big facts about growth. *Harvard Kennedy School of Government*.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reed, L. and Jeremiah, J. (2017). Educational outcomes: Adversity and resilience. *Developments in Business Simulation and Experiential Learning*, 44(1):252–256.

Robinson, D. and Hayes, A. (2019). *broom: Convert Statistical Analysis Objects into Tidy Tibbles*. R package version 0.5.2.

Romer, P. (1990). Capital, labor, and productivity. *Brookings Papers on Economic Activity*, 21(1990 Microeconomics):337–367.

Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of Political Economy*, 94(5):1002–1037.

Rudis, B., Bolker, B., Schulz, J., Kothari, A., and Sidi, J. (2019). *ggalt: Extra Coordinate Systems, 'Geoms', Statistical Transformations, Scales and Fonts for 'ggplot2'*. R package version 0.6.2.

Salih, A. B. (2016). Improving education in sudan to empower future generations. https://www.worldbank.org/en/news/feature/2016/01/07/improving-education-in-sudan-to-empower-future-generations.

Samuelson, P. A., Nordhaus, W. D., and Chaudhuri, S. (2011). *Macroeconomics*. McGraw Hill Higher Education.

Schultz, T. (1961). Investment in human capital. *The American Economic Review*, 51(1):1–17.

Scotland, J. (2012). Exploring the philosophical underpinnings of research: Relating ontology and epistemology to the methodology and methods of the scientific, interpretive, and critical research paradigms. *English Language Teaching*, 5(9).

Sherriff, L. (2015). Ernst & young removes university degree classification from entry criteria as there's 'no evidence' it equals success.

Siddig, E.-F. A. and Hassan, F. M. (2016). The sudan interim poverty reduction strategy paper status report : a joint world bank group and sudan's ministry of finance and economic planning assessment (english). *The World Bank Group*, page 1–112.

Slowikowski, K. (2019). *ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'*. R package version 0.8.1.

Smith, A. (1786). *An inquiry into the nature and causes of the wealth of nations*. Cosmio.

Solow, R. M. (1956). A contribution to the theory of economic growth. *The Quarterly Journal of Economics*, 70(1):65–94.

South, A. (2017). *rnaturalearth: World Map Data from Natural Earth*. R package version 0.1.0.

Sustainable Development Goals (2015). Goal 4: Quality education. https://www.globalgoals.org/4-quality-education.

Swan, T. W. (1956). ECONOMIC GROWTH and CAPITAL ACCUMULATION. *Economic Record*, 32(2):334–361.

Tan, E. (2014). Human Capital Theory: A Holistic Criticism. *Review of Educational Research*, 84(3):411–445.

Taylor, G. (2018). An introduction to latex [class presentation]. *Oxford Internet Institute, University of Oxford*, page 1–108.

Taylor, S. and Yu, D. (2009). The importance of socioeconomic status in determining educational achievement in south africa. *Bureau for Economic Research (University of Stellenbosch)*, 01(09):1–77.

Tennekes, M. (2018). tmap: Thematic maps in R. *Journal of Statistical Software*, 84(6):1–39.

The Fund for Peace (2019). Global data. https://fragilestatesindex.org/data/.

The Nobel Prize (1987). Press release 1987. https://www.nobelprize.org/prizes/economic-sciences/1987/press-release/.

Trainor, A. A. and Graue, E. (2014). Evaluating rigor in qualitative methodology and research dissemination. *Remedial and Special Education*, 35(5):267–274.

Transparency International (2018). Corruption perceptions index 2018. https://www.transparency.org/cpi2018.

Trapletti, A. and Hornik, K. (2019). *tseries: Time Series Analysis and Computational Finance*. R package version 0.10-47.

UNESCO (2004). Efa global monitoring report: The quality imperative. regional overview. *the United Nations Educational, Scientific and Cultural Organization*, page 1–428.

UNESCO (2018). *Sudan education policy review: Paving the road to 2030*. ED-2018/WS/9-CLD 706.18. Section of Education Policy (UNESCO).

UNICEF (2016). Education. https://www.unicef.org/infobycountry/files/Unicef_Sudan_EDUCATION_PROGRAMME_FINAL_(032016).pdf.

United States Department of the Treasury (2019). Sudan and darfur sanctions. https://www.treasury.gov/resource-center/sanctions/Programs/pages/sudan.aspx.

Walsh, D. (2019). Sudan power-sharing deal reached by military and civilian leaders. https://www.nytimes.com/2019/07/04/world/africa/sudan-power-sharing-deal.html.

Watts, A. (2017). International surveys TIMSS, PISA, PIRLS Issue Brief No. 6920565690. *Cambridge Assessment International Education*, page 1–5.

Whelan, K. (2015). Advanced macroeconomics. *University College Dublin*, page 1–29.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, H. (2017). *tidyverse: Easily Install and Load the 'Tidyverse'*. R package version 1.2.1.

Wickham, H. (2019). *rvest: Easily Harvest (Scrape) Web Pages*. R package version 0.3.4.

Wickham, H. and Bryan, J. (2019). *readxl: Read Excel Files*. R package version 1.3.1.

Wickham, H., François, R., Henry, L., and Müller, K. (2019). *dplyr: A Grammar of Data Manipulation*. R package version 0.8.2.

Wilson, T. (2019). Sudan internet blackout forces battered protesters to rethink. https://www.ft.com/content/b1848126-8c0f-11e9-a1c1-51bf8f989972.

World Bank (2012). *The Status of the Education Sector in Sudan (Africa Human Development Series)*. The World Bank Group.

World Bank (2017a). The human capital index. http://www.worldbank.org/en/publication/human-capital.

World Bank (2017b). Skills development. http://www.worldbank.org/en/topic/skillsdevelopment.

World Bank (2018). The world bank in sudan. https://www.worldbank.org/en/country/sudan/overview#3.

World Bank (2019a). Sudan basic education recovery project. http://projects.worldbank.org/P128644/sudan-basic-education-recovery-project?lang=en&tab=overview.

World Bank (2019b). Sudan country profile. data retrieved from World Development Indicators, https://databank.worldbank.org/views/reports/reportwidget.aspx?Report_Name=CountryProfile&Id=b450fd57&tbar=y&dd=y&inf=n&zm=n&country=SDN.

Xiao, N. (2018). *ggsci: Scientific Journal and Sci-Fi Themed Color Palettes for 'ggplot2'*. R package version 2.9.

Zdanowicz, J. S. (2009). Trade-based money laundering and terrorist financing. *Review of Law & Economics*, 5(2).

Zeileis, A. (2017). *pwt9: Penn World Table (Version 9.x)*. R package version 9.0-0.

Appendix A

Further Descriptions & Analyses

Chapter 1

Section 1.3.1

Comparing Major International Student Assessments: PISA–TIMMS–PIRLS

	PISA	**TIMMS**	**PIRLS**
Name	Program for International Student Assessment	Trends in International Mathematics and Science Study	Progress in International Reading Literacy Study
Assessment Type	Reading, mathematics, science, problem solving	Mathematics and science	Reading
Age Cohort	15	10 and 14	10
Grade in School	Grade 9 (UK Year 10)	Grade 4 and Grade 8 (UK Years 5 and 9)	Grade 4 (UK Year 5)
Frequency	3 years(since 2000)	4 years(since 1995)	5 years(since 2001)

Assessment Purpose	Evaluates whether students at the end of compulsory education can apply their knowledge to real-life situations and be equipped for society	Measures maths and science achievement. Describes students' attitudes, curriculum, teachers'training, classroom activities	Measures trends in comprehension for reading. Investigates the experiences young children have at home and school in learning to read
Assessment Focus	Skills	Curriculum	Curriculum
Organization	OECD	IEA	IEA
Countries	72 countries and economies	57 countries and 7 bench-marking entities	50 countries and 11 bench-marking entities
Length of Examination	120 minutes; 35 minute background questionnaire	72 minutes at Grade 4; 90 minutes at grade 8; 15 minute background questionnaire	80 minutes; 15 minute background questionnaire
Sample Size	5,000+ learners per country/jurisdiction	4,000+ learners per country/jurisdiction	3,500–4,000 learners per country/ jurisdiction

Source: (Watts, 2017, p 2)

Figure 1.4

In Hanushek and Woessmann (2015, p. 4) Region abbreviations are described as: Central Europe (C-EUR); Commonwealth OECD members (COMM); East Asia (E-ASIA); Latin America (LATAM); Middle East and North Africa(MENA);Northern Europe (N-EUR);Southern Europe(S-EUR); sub-Saharan Africa (SSAFR). Commonwealth

OECD states are Australia, Canada, New Zealand, and the US.

Barro and Sala-i Martin (2004)'s Basic Cross–Country Growth Regressions

(1) Explanatory Variable	(2) Coefficient	(3) Low-Income Sample	(4) High-Income Sample	(5) *p value*[a]	(6) 5-Year Intervals
Log of GDP per capita	-0.248(0.0029)	-0.021(0.005)	-0.032(0.005)	0.12	-0.0237(0.0029)
Male Upper-level schooling	0.004(0.002)	0.006(0.005)	0.002(0.002)	0.44	0.002(0.002)
1/(life expectancy at age 1)	-5.04(0.86)	-5.13(1.18)	-1.28(1.44)	0.040	-4.91(0.90)
Log of total fertility rate	-0.012(0.005)	-0.021(0.012)	-0.021(0.005)	0.99	-0.016(0.0048)
Government consumption ratio	-0.06 (0.02)	-0.10 (0.03)	-0.000 (0.03)	0.021	-0.066(0.021)
Rule of law	0.019(0.006)	0.024(0.01)	0.022(0.006)	0.90	0.017(0.006)
Democracy	0.08(0.03)	0.044(0.05)	0.101(0.04)	0.32[b]	0.032(0.017)
Democracy squared	-0.07(0.03)	-0.05(0.05)	-0.08(0.03)	0.67	-0.03(0.02)
Openness ratio	0.00(0.005)	0.017(0.011)	0.006(0.0046)	0.38	0.01(0.0043)
Change in trade	0.13(0.05)	0.18(0.08)	0.04(0.07)	0.16	0.03(0.021)
Investment ratio	0.08(0.024)	0.11(0.04)	0.08(0.04)	0.46	0.058(0.022)
Inflation rate	-0.019 0.010)	-0.019(0.012)	-0.019(0.009)	0.99	-0.031(0.007)
Constant	0.30(0.034)	0.29(0.05)	0.30(0.052)	0.99[c]	0.31(0.035)
Dummy, 1975–85	-0.008(0.003)	-0.01(0.004)	-0.01(0.0032)	0.81	[d]
Dummy, 1985–95	-0.013(0.003)	-0.019(0.005)	-0.005(0.004)	0.031	
Number of observations	72,86,83	26,38,33	46,48,50		72,79,86,84, 79,80,60
R-squared	.60,.49,.51	.78,.53,.65	.56,.56,.40		.40,.26,.27,.31, .46,.19,.04

Notes Estimation is by three-stage OLS. In column 2 the DVs are the GDP per capita growth rates. Columns 3 and 4 separate the samples into countries with levels of GDP per capita below and above the median (for 1960, 1970, and 1980). Column 6 uses equations for economic growth for seven five-year periods, 1965–70, . . . , 1995–2000.

[a] *p*values refer to the hypothesis that coefficients are the same for the two income groups

[b] p value for democracy and democracy-squared is 0.022

[c] p value for the constant and two-time dummies is 0.010

[d] The time dummies at the 5-year intervals are -0.0014(0.0040) for 1970–75, -0.0000(0.0040) for 1975–80, -0.0180(0.0040) for 1980–85, -0.0112(0.0037) for 1985–90, -0.0184(0.0045) for 1990–95, and -0.0165(0.0042) for 1995–2000.

Source (Barro and Sala-i Martin, 2004, p 512)

Chapter 2

Section 2.1

Charges Against Al-Bashir

As a result of the atrocious events in Darfur and Kassala regions in Sudan, the International Criminal Court issued two arrests ordered for the 7^{th} President of Sudan, Omar Al-Bashir in 2009 and 2010 for the following charges (Coalition for the International Criminal Court, 2009, para 6 and 7):

1. Al-Bashir is suspected of committing, as an indirect perpetrator, five counts of crimes against humanity (murder, extermination, forcible transfer, torture, and rape), two counts of war crimes (intentionally directing attacks against a civilian population and pillaging) and three counts of genocide during the alleged counter-insurgency campaign [para 6].
2. The three counts of genocide include genocide by killing; genocide by causing serious bodily or mental harm; and genocide by deliberately inflicting conditions of life calculated to bring about physical destruction [para 7].

Section 2.2.1

Components of the 2019 EFI

- *The Rule of Law*: This measure consists of property rights, judicial effectiveness, and government integrity. The Index states that the capital city mainly controls power and resources; the enforcement of property rights is lackluster, and the judiciary is far from independent. As also stated in (Heritage Foundation, 2019), Sudan is mentioned as one of the most corrupt nations in the world as it was ranked the 172^{nd} out of 180 countries on *Corruption Perceptions Index* in 2018 (Transparency International, 2018).

- *Regulatory Inefficiency*: This measure is made of business, labor, and monetary freedom. The Index states that "a large informal economy remains trapped by business regulations that inhibit registration and a rigid labor market that discourages formal hiring" [p.391]. Furthermore, Khartoum's decision to remove fuel subsidies and wheat subsidies hiked up the inflation and led to long-lasting street protests, which ended up in Al-Bashir's removal.

- *Open Markets*: This measure shows how free a country's trade, investment, and financial freedom. Sudan scored 5 out of 100 on investment freedom, arguably one of the lowest in the world. The Index states that this is due to low access to credit and the fact that only 17% of the Sudanese have access to banking services. One of the concepts that contributed to such a low percentage of access to banking services could be the comprehensive US sanctions on Sudan, which were lifted in 2017 (Central Intelligence Agency, 2019). Given the majority of global banking services are developed by the US and the countries that are US allies such as the UK, Switzerland, and Germany, these countries also follow US sanctions (United States Department of the Treasury, 2019). Henceforward, these sanctions are one of the reasons why the economy is mostly informal, and daily transactions occur on a cash basis, which makes the supervision of accounting, money flows and remittances tough and opens room for illegal activities such as money laundering (Zdanowicz, 2009).

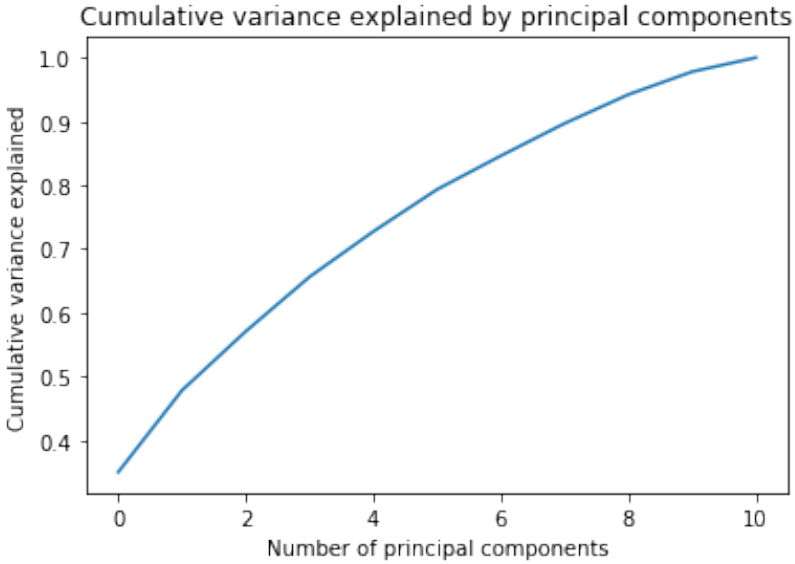
Chapter 3

Principal Component Analysis: Section 3.2.2

Data: Johnson (2018); Software: Phyton Programming Language[1] (McKinney, 2010).

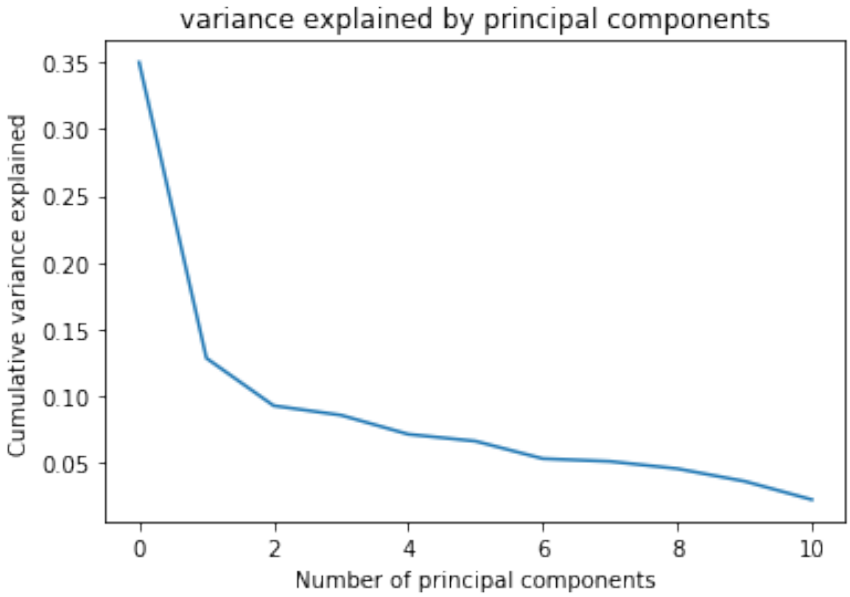
a. PCA Components

Variable	Explained Variance Ratio
Electricity Power	0.4620428
TV	0.4627492
Refrigerator	0.46460461
Phone	0.14499398
Motorbike	0.12196736
Car	0.24548517
Study Table	0.29289631
Private Room	0.14383898
Place to Study	0.24787569
Computer	0.15650434
Books	0.26066466
Overall Array	0.35016153

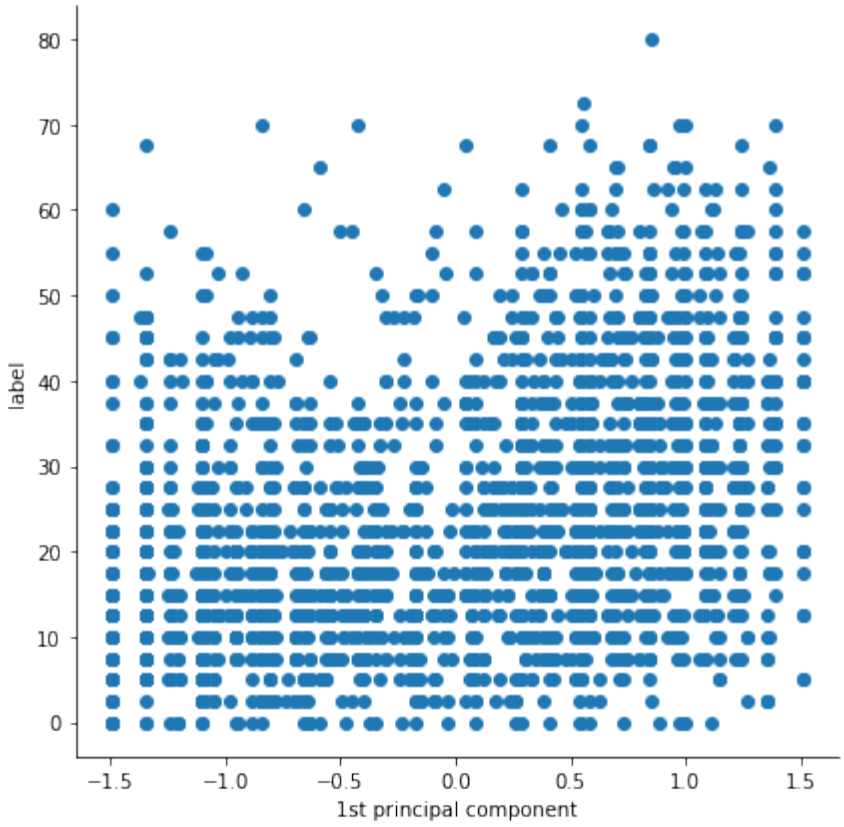


b. Cumulative Variance Explained By Principle Components

[1]See the code in Appendix **B**



c. Variance Explained by Principal Components



d. Variance Explained by the First Principal Component

Chapter 4

Section 4.2

Regression Assumptions Checks

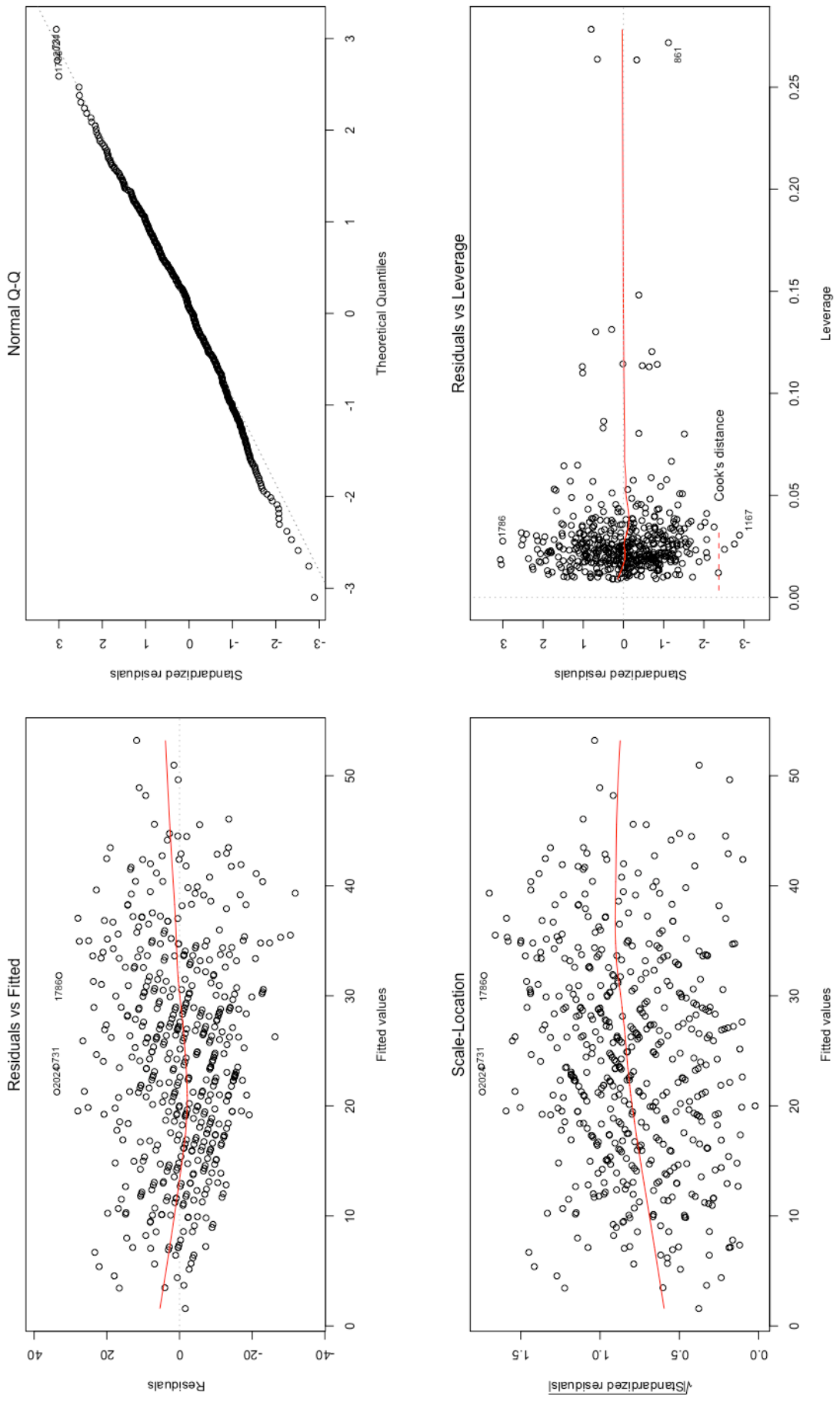
a. AIC

	Degrees of Freedom	AIC
Model 1	3	4527.049
Model 2	4	4499.807
Model 3	8	4495.823
Model 4	10	4122.595
Model 5	16	3989.784

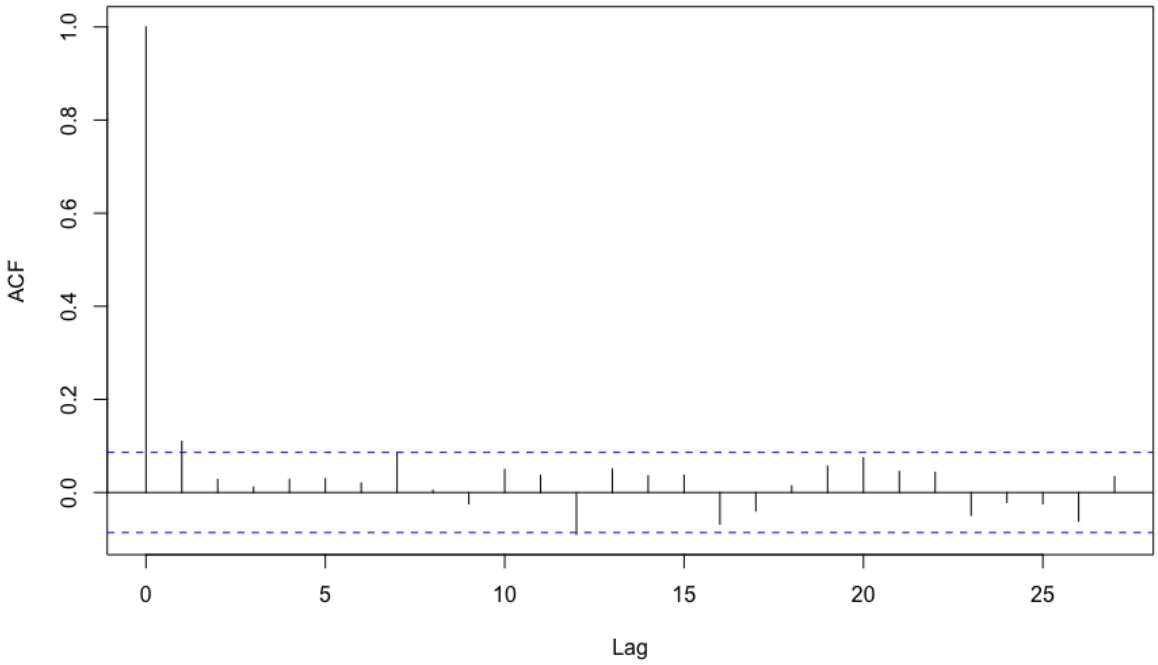
- **Runs Test:** Standardized Runs Statistic = -0.96756, p-value = 0.3333
  - With a p-value = 0.3333, the null hypothesis (residuals are random) cannot be rejected. Therefore it can safely be assumed that there is no autocorrelation in residuals.
- **Durbin Watson Test Value :** 1.778879
- **Mean of Residuals:** -6.785347 e^{-17}

b. VIF

	GVIF	Degrees of Freedom	$GVIF^{1/(2*Df)}$
Math Achievement	1.086371	1	1.042291
SAI	1.692601	1	1.301000
School Type	1.244226	3	1.037090
Gender	1.027911	1	1.013860
Father work status	1.093761	1	1.045830
Mother work status	1.139969	1	1.067693
Father education type	2.960360	3	1.198278
Mother education type	2.964246	3	1.198540



c. Regression Plots



d. ACF Plot

Chapter 4

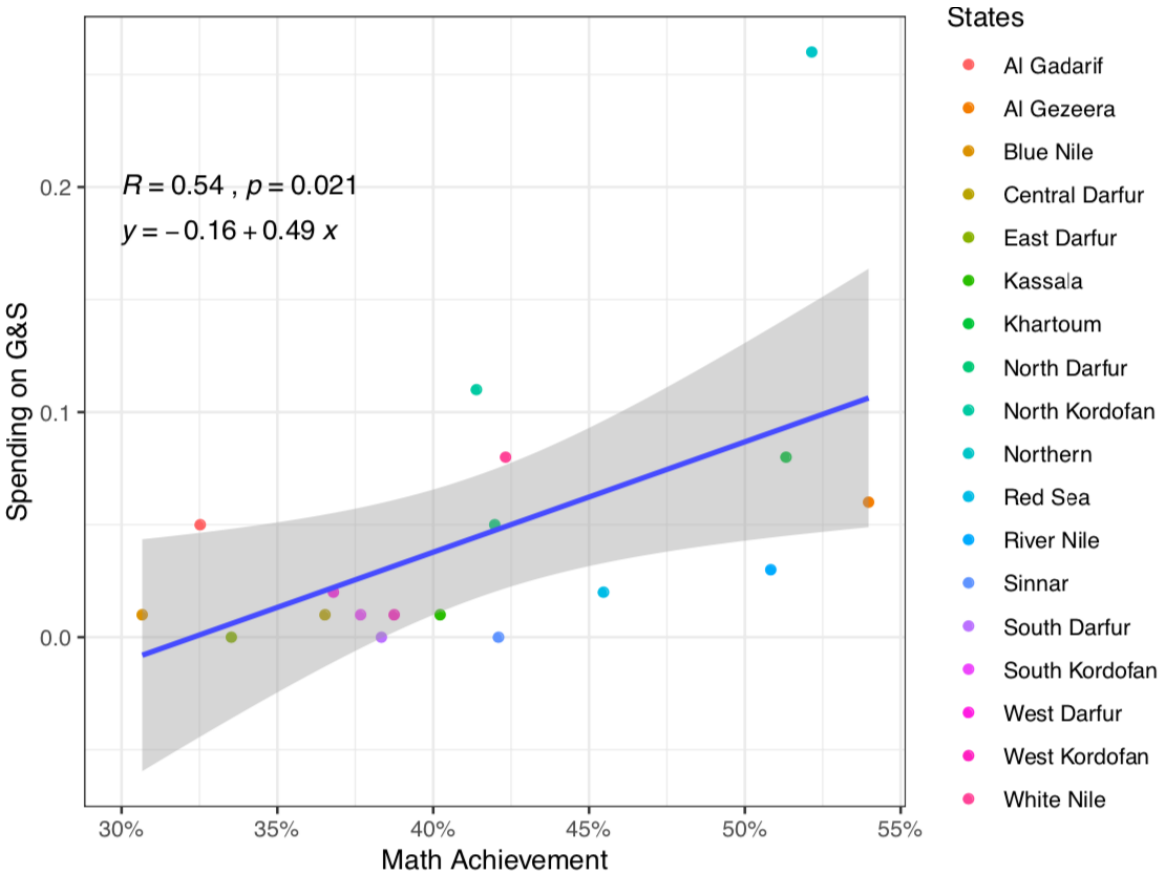
Section 4.1

Additional Correlation Plots

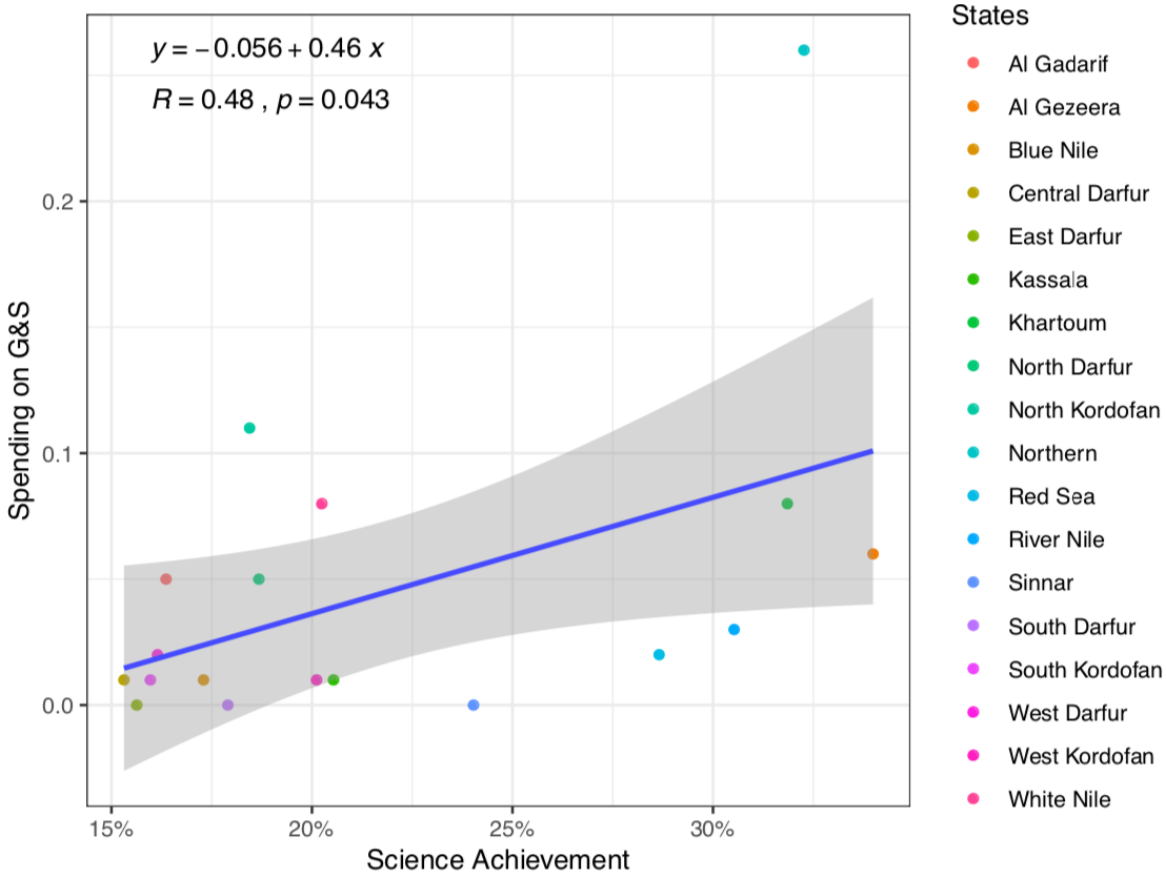
As per mentioned in Section **4.1**, additional correlation plots were created to show the relationship between Science& Math Achievement and the components of state-level educational spending in Sudan; and not included given the word-limit restriction.

As the plots show below, spending on goods & services formed a positive relationship with both science & math scores, significant at $p<0.05$. Spending on goods & services was able to explain on average 51 % of the variation in science & math scores; and a percentage point increase in spending on goods & services indicated on average 0.46 percentage point increase in science & math within the state-level data where sample size is 18.

On the other hand, spending on teachers formed a negative relationship with both science & math scores, significant at $p<0.05$. Spending on teachers was able to explain on average 49 % of the variation in science & math scores; and a percentage point increase in teachers indicated on average 0.44 percentage point decrease in science & math within the state-level data where sample size is 18.

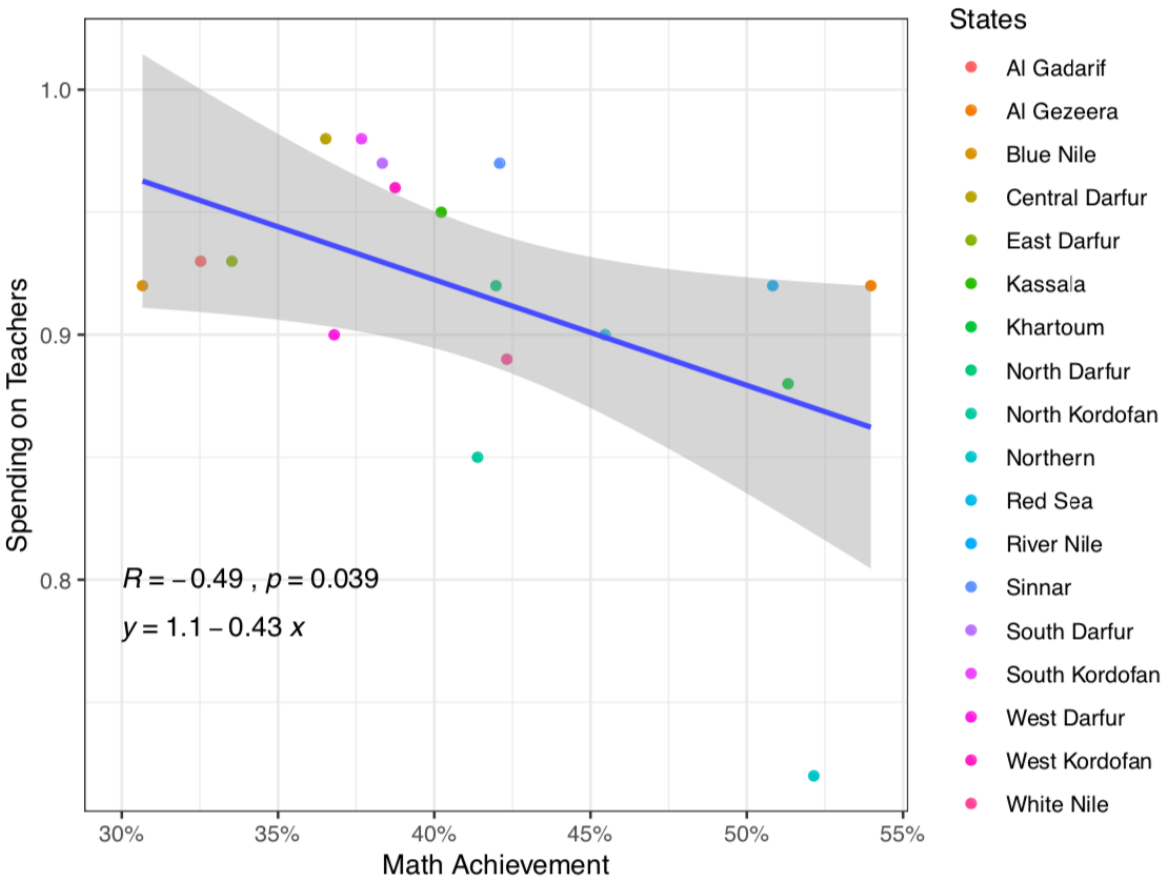


Correlation between Math Achievement and Spending on Goods & Services in Sudan

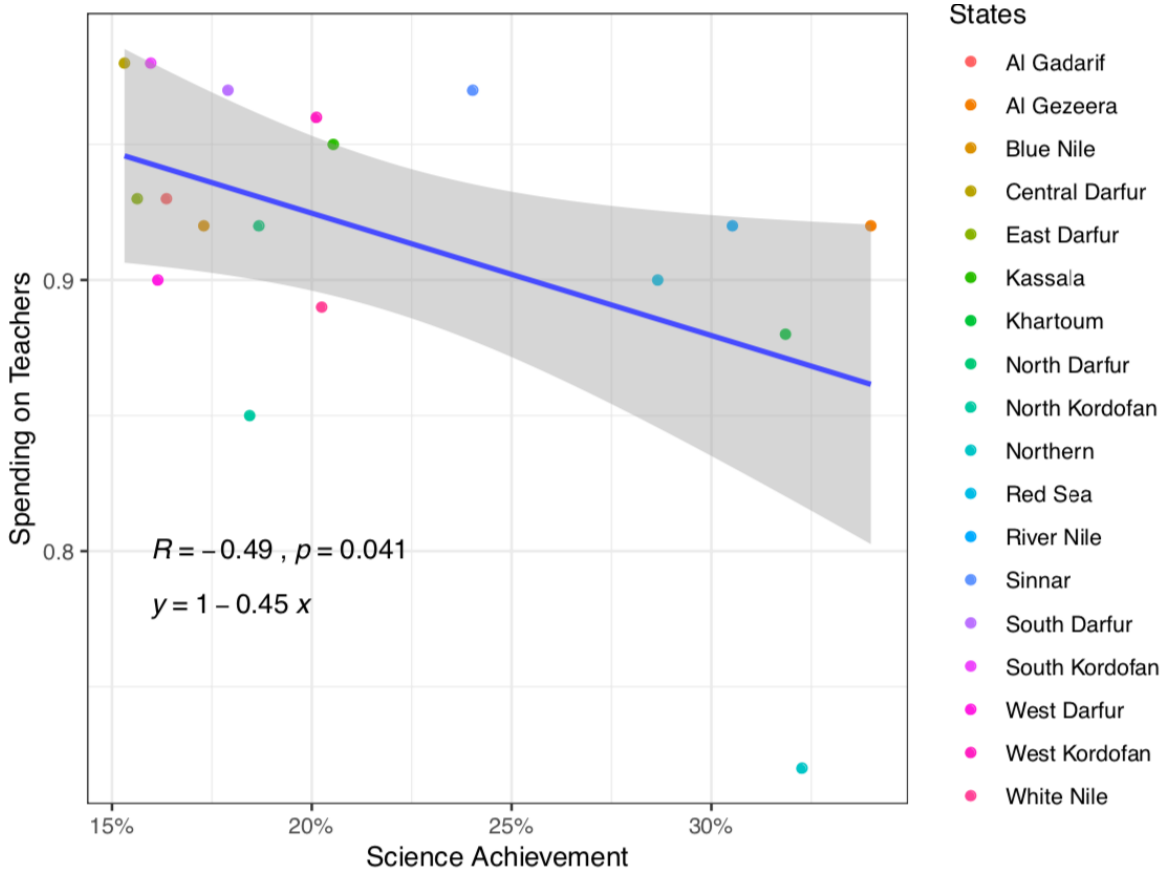


Correlation between Science Achievement and Spending on Goods & Services in Sudan

Correlations between Science& Math Achievement and Spending on Goods & Services in Sudan. Achievement Data: Johnson (2018); Spending Data: Siddig and Hassan (2016)

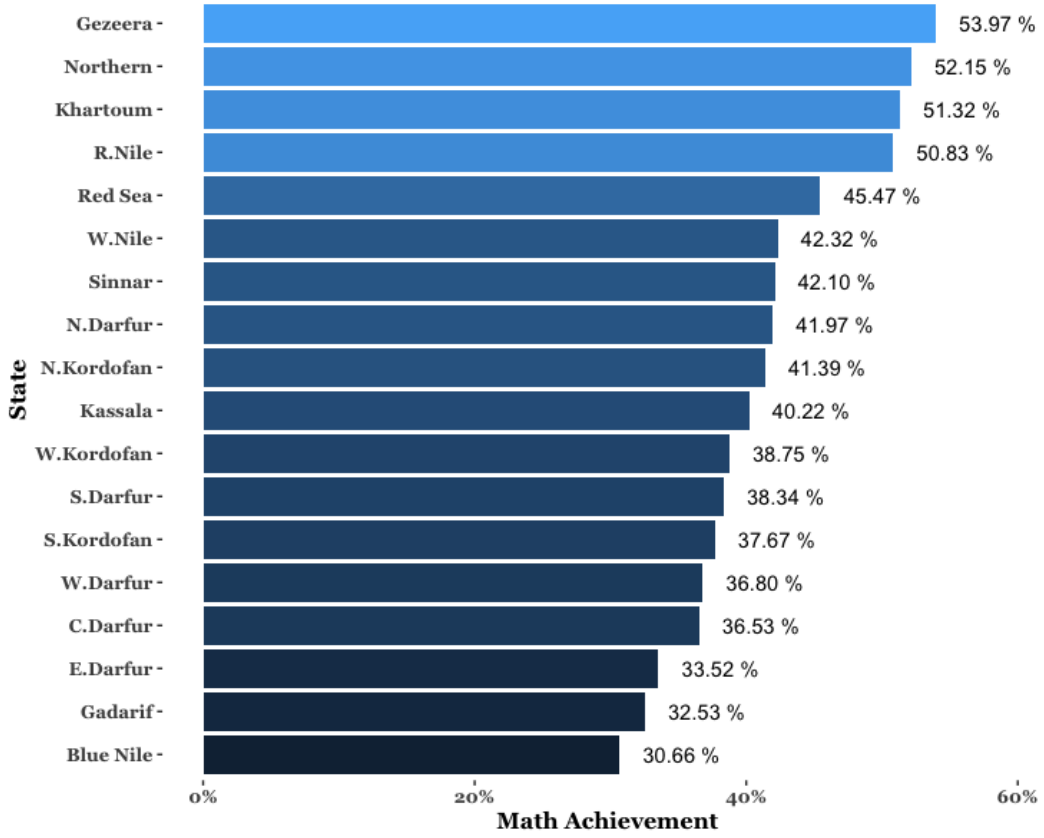


Correlation between Math Achievement and Spending on Teachers in Sudan

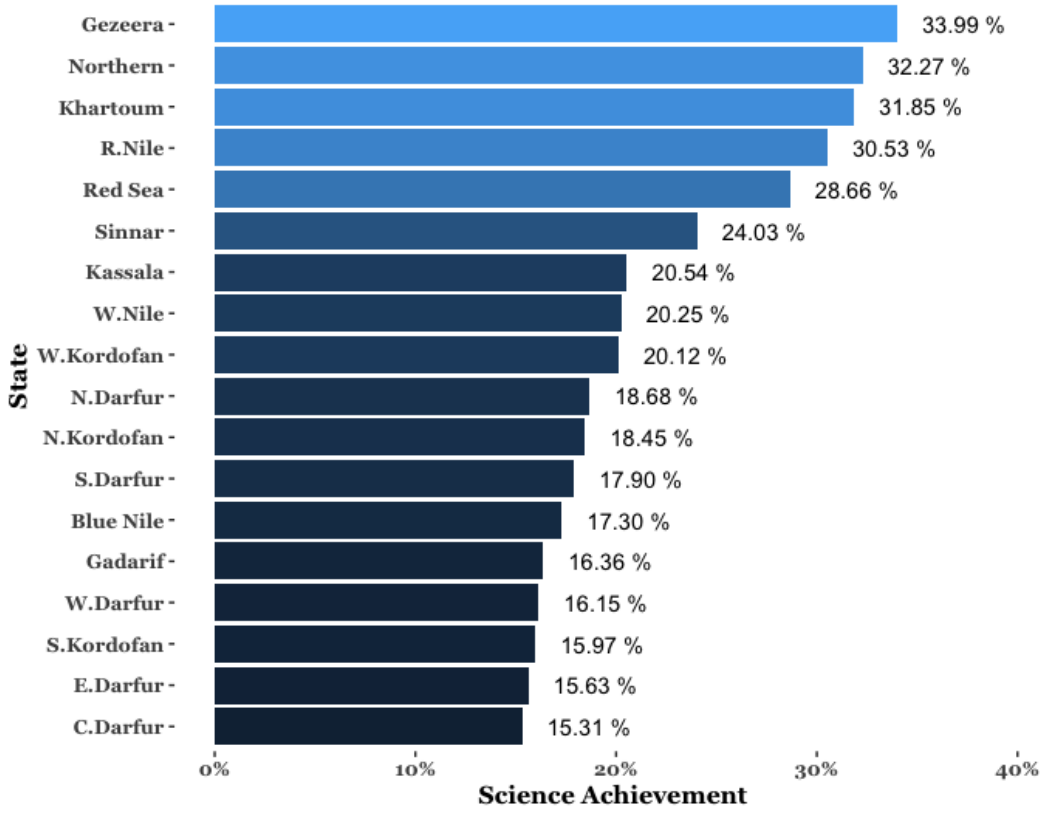


Correlation between Science Achievement and Spending on Teachers in Sudan

Correlations between Science& Math Achievement and Spending on Teachers in Sudan. Achievement Data: Johnson (2018); Spending Data: Siddig and Hassan (2016)



(a) Math Achievement by State



(b) Science Achievement by State

2018 NLA Achievement in Science and Math, by State. Data: Johnson (2018)

Appendix B

Codes

Codes for Creating Plots

Plots in Section 2.1

Africa Maps with World Development Indicators in Section 2.1

The following code was written on RStudio software [1] (R Core Team, 2019) to create the plots in Figure 2.1. The basis of the following code was developed by Lovelace et al. (2019). The code was then altered and tweaked for creating the plots in thesis. The following is a list of R Packages used:

1. raster: (Hijmans, 2019)
2. sf: (Pebesma, 2018)
3. tmap: (Tennekes, 2018)
4. dplyr: (Wickham et al., 2019)
5. spData: (Bivand et al., 2019)
6. leaflet: (Cheng et al., 2018)
7. cartogram: (Jeworutzki, 2018)
8. tidyverse: (Wickham, 2017)

[1] ***** in "setwd" is used to anonymize the author of the thesis

9. remotes: (Hester et al., 2019)
10. RColorBrewer: (Neuwirth, 2014)

The code written for the plots:

```
setwd("/Users/***********/Desktop/graphs")
install.packages(c("sf", "raster", "tidyverse", "spData", "tmap",
"leaflet", "dplyr","cartogram", "remotes"))

 mapsafrica = world %>%
  filter(continent == "Africa", !is.na(iso_a2)) %>%
  left_join(worldbank_df, by = "iso_a2") %>%
  dplyr::select(name, subregion, gdpPercap, HDI, pop_growth) %>%
  st_transform("+proj=aea +lat_1=20 +lat_2=-23 +lat_0=0 +lon_0=25")

map1 = tm_shape(mapsafrica) +
  tm_polygons(col = "HDI", title = "HDI",
    breaks = c(0, 0.55, 0.7, 0.8),
    labels = c("Low", "Medium", "High"),
    palette = "Blues")

map2 = tm_shape(mapsafrica) + tm_polygons(col = "subregion",
title = "SUBREGION", palette = "Accent")

map3 = tm_shape(mapsafrica) + tm_polygons("pop_growth",
title = "POPULATION GROWTH", palette="Spectral")

tmap_arrange(map1, map2,map3)
```

2019 Fragile States Index: Section 2.1

a. Africa, Figure 2.3

The following code was written on RStudio software (R Core Team, 2019) to create the plots in Figure 2.3. The basis of the following code was developed by Luginbühl (2018). The code was then altered and tweaked for creating the plots in thesis

In addition to the R Packages mentioned above, the following packages were used to create the plots:

1. rvest: (Wickham, 2019)
2. ggalt: (Rudis et al., 2019)
3. readxl: (Wickham and Bryan, 2019)
4. ggplot2: (Wickham, 2016)
5. ggthemes: (Arnold, 2019)
6. countrycode: (Arel-Bundock et al., 2018)
7. rnaturalearth: (South, 2017)

The following is the code written for the plots:

```
setwd("/Users/***********/Desktop/graphs")

library(rvest)
library(readxl)
library(tidyverse)
library(RColorBrewer)

install.packages("ggalt")
remotes::install_github("hrbrmstr/ggalt")

install.packages("ggthemes")
remotes::install_github("jrnold/ggthemes")

library(ggalt)
library(ggthemes)
```

```
index_2019 <- read_excel(path = "fsi-2019.xlsx")
index_2019 %>% select(Country, Year, Rank, Total)

map_breaks = c(0, 30, 40, 50, 60, 70, 80, 90,
100, 110, 120)

label_breaks = c("0-30", "30-40", "40-50", "50-60",
"60-70", "70-80", "80-90", "90-100", "100-110", "110-120")

index_2019$Classes <- cut(index_2019$Total,
                          breaks = map_breaks,
                          labels = label_breaks)

world <- ggplot2::map_data("world") %>%
  filter(region!= "Antartica")

dplyr::setdiff(index_2019$Country, world$region)

######FOR AFRICA######

world2 <- ggplot2::map_data("world") %>%
  filter(region== c("Sudan","Algeria", "Egypt", "Libya",
  "Morocco","Tunisia", "Western Sahara",
  "South Africa", "Madagascar"))

ggplot() +
  geom_map(data = world2, map = world2,
           aes(x = long, y = lat, map_id = region),
           col = "grey", fill= "white", size = 0.1) +
  geom_map(data = index_2019, map = world,
           aes(map_id = Country, fill = Classes)) +
  ggalt::coord_proj("+proj=wintri") +
  scale_fill_brewer(NULL, palette = "RdYlGn", direction = -1) +
  ggthemes::theme_map() +
  theme(plot.title = element_text(face = "bold"))
```

b. World: Figure 2.4

Figure 2.4 was built in conjunction with figure 2.3. In addition to the code, citations and packages above, the following code was written produce the plot:

```
library(rnaturalearth)
library(countrycode)
library(sf)

earth_df <- ne_countries(scale = 50, returnclass = "sf") %>%
  select(admin, geometry, iso_a2)

index_2019$iso2 <- countrycode(index_2019$Country, "country.name",
  "iso2c")

map_data_df <- inner_join(earth_df, index_2019, by =
  c("iso_a2" = "iso2")) %>%
  select(Country, Total, Classes, geometry) %>% sf::st_as_sf()

map_data_df %>%
  ggplot() +
  geom_sf(aes(fill = Classes), size = 0.1, color = "white") +
  coord_sf(crs = "+proj=moll") +
  scale_fill_brewer(NULL,
                    palette = "RdYlGn",
                    direction = -1,
                    guide = guide_legend(
                      title = "More to less stable",
                      title.hjust = 0.5,
                      direction = "horizontal",
                      nrow = 1,
                      keyheight = unit(3, units = "mm"),
                      keywidth = unit(3, units = "mm"),
                      label.position = "right",
                      title.position = "top")) +
  theme_minimal() +
  theme(legend.direction = "horizontal",
        legend.position = "bottom")
```

Percentage of Children out of Primary School in Sudan, by State (2016): Section 2.3.4

The following code was written on RStudio software (R Core Team, 2019) to create the plots in Figure 2.3. The code below is unique to the author of this thesis, however the author acknowledges important learnings and direction given by Healy (2019) in his book *Data Visualizaton: A Practical Introduction* to make the creation of plot better in terms of visualization, and the code written more efficient. Data used to create this plot was taken from UNICEF (2016). In addition to the packages cited above, the following packages were used to create plot :

1. ggsci: (Xiao, 2018)
2. broom: (Robinson and Hayes, 2019)
3. ggpubr: (Kassambara, 2018)
4. ggrepel: (Slowikowski, 2019)
5. gridextra: (Auguie, 2017)
6. extrafont: (Chang, 2014)
7. extrafontdb: (Chang, 2012)

The following is the code written for the plots:

```
setwd("/Users/**************/Desktop")

sudan_state<- read_excel(path=
"/Users/*********/Desktop/sudan_state_level.xlsx")

library("ggsci")
library("ggplot2")
library("gridExtra")
library("ggpubr")
library("ggrepel")
library("tidyverse")
library("broom")
library("ggthemes")
```

```
install.packages("extrafontdb")
library(extrafontdb)

install.packages("extrafont")
library(extrafont)

font_import()
fonts()

p00 <- ggplot(data = sudan_state,
aes (x=reorder(state,out_of_schools), y=out_of_schools,
fill=out_of_schools))

p00+geom_col(position="dodge2")+ guides(fill=FALSE)+
coord_flip()+ scale_y_continuous(labels =
    scales::percent_format(accuracy = 1L),
    limits=c(0,0.5))+ labs(
    x= "State",
    y= "Percentage of Children out of Primary School ",
    fill= "out_of_schools") +
    geom_text(data = sudan_state,
    aes(x=reorder(state,out_of_schools),
    y=out_of_schools,
    label = printf("%.02f %%", out_of_schools*100)),
    position = position_dodge(
    width = 1),
    hjust=-0.3) +
    theme_tufte()+
    theme(text =  element_text(
    size=12, family="Georgia"))+
    scale_color_hc()
```

Section 4.1

Figure 4.1

No packages in addition to the packages cited above were used to create the Figure **4.1**

```
setwd("/Users/********/Desktop")
sudan_state <- read_excel(path= "/Users/*********/
Desktop/sudan_state_level.xlsx")

font_import()
fonts()

pdf("cor_plot.pdf", width = par('din')[1],
    height = par('din')[2], onefile=T)

p1 <- ggplot(data = sudan_state, mapping = aes
(x=science,y=pop_under_pov))

p1+ geom_point(mapping= aes(color=state))+ geom_smooth
    (method="lm")+ theme_bw() + labs(
        y ="Population living below poverty line",
        x="Science Achievement",
        color= "States",
        fill="States")+
        scale_x_continuous( labels =
        scales::percent_format(accuracy = 1L))+
        stat_cor(label.x = 0.28,
        label.y = 0.62) + stat_regline_equation(
        label.x = 0.28, label.y = 0.65)

p2 <- ggplot(data = sudan_state, mapping = aes
    (x=science,y=pop_under_pov))

p2+ geom_point(mapping= aes(color=state))+ geom_smooth
    (method="lm")+ theme_bw() + labs(
        y ="Population living below poverty line",
```

```
        x="Science Achievement",
        color= "States",
        fill="States")+
        scale_x_continuous(labels =
        scales::percent_format(accuracy = 1L))+
        stat_cor(label.x = 0.28,
        label.y = 0.62) +
        stat_regline_equation(
        label.x = 0.28, label.y = 0.65)
dev.off()
```

Principal Component Analysis: Section 3.2.2

Data: Johnson (2018); Software: Phyton Programming Language (McKinney, 2010).

```
In [30]: import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt

In [57]: data = pd.read_csv
('clean_data_merged copy 2.csv')

In [57]: for col in data.drop(columns='agg_percentage_Science')
.columns: data[col] = data[col].apply
(lambda x: 1 if x=='Yes' else 0)

In [6]: from sklearn.decomposition import PCA

In [60]: pca = PCA()
         pca.fit(data.drop(columns='agg_percentage_Science'))
```

Out [**60**]: PCA(copy=True, iterated_power='auto',
n_components = None, random_state = None, svd_solver ='auto',
tol = 0.0, whiten=False)

```
In [61]: plt.plot(pca.explained_variance_ratio_.cumsum())
         plt.title('Cumulative variance explained
```

```
        by principal components')
        plt.xlabel('Number of principal components')
        plt.ylabel('Cumulative variance explained')
        plt.show()
```

Out [ ]: See Appendix **A**: Principal Component Analysis: Section 3.2.2 (Figure b.)

```
In [61]: plt.plot(pca.explained_variance_ratio_)
         plt.title('variance explained by
         principal components')
         plt.xlabel('Number of principal components')
         plt.ylabel('Cumulative variance explained')
         plt.show()
```

Out [ ]: See Appendix **A**: Principal Component Analysis: Section 3.2.2 (Figure c.)

```
In [62]: data.head(3)
In [70]: pca2 = PCA(n_components=1)
         df_pca2 = pca2.fit_transform
         (data.drop(columns='agg_percentage_Science'))
         display(pca2.components_)
         print(pca2.explained_variance_ratio_)
```

array ([[0.4620428 , 0.4627492 , 0.46460461, 0.14499398,
0.12196736, 0.24548517, 0.29289631, 0.14383898,
0.24787569, 0.15650434, 0.26066466]])

[0.35016153]

See Appendix **A**: Principal Component Analysis: Section 3.2.2 (Table a.)

```
In [65]: np.array(df_pca2).shape
```

Out [**65**]: (2172, 1)

```
In [65]: data['agg_percentage_Science'].shape
```

Out [**65**]: (15222,)

```
In [65]: data['agg_percentage_Science'].shape
```

```
  In [65]:  import seaborn as sn
# attach label for each 2-d data point in NLA data
    pca2_data = np.concatenate((np.array(df_pca2),

    np.array(data['agg_percentage_Science'])
    .reshape(2172,1)),axis=1)

# creating a new data frame to plot the result NLA data
    pca_df = pd.DataFrame(data=pca2_data,
    columns=("1st principal component", "label"))
    sn.FacetGrid(pca_df, size=6).map(plt.scatter,
    '1st principal component', 'label')
    .add_legend()

# plt.title('2D plot of the NLA data with the first 2 principal components')
plt.show()
```

Out [ ]: See Appendix **A**: Principal Component Analysis: Section 3.2.2 (Figure d.)

Data Merging

Data: Johnson (2018); Software: Phyton Programming Language (McKinney, 2010)

```
#!/usr/bin/env python2
# -*- coding: utf-8 -*-

import pandas as pd

df_science=pd.read_csv('science grade six.csv')
science_col=['Q_'+str(k) for k in range(1,10)]
+['agg_score','agg_percentage']
df_science=df_science[science_col]

df_science=df_science.rename(index=str,
columns={"id_1": "id_1_science",
"id_2": "id_2_science"})
df_science=df_science.drop_duplicates
(subset='Q_5',keep=False)
```

```
df_science=df_science.set_index('Q_5')


df_math=pd.read_csv('math grade six.csv')
math_col=['Q_'+str(k) for k in range(1,10)]
+['agg_score','agg_percentage']
df_math=df_math[math_col]

df_math=df_math.rename(index=str,
columns={"agg_score": "agg_score_math",
"agg_percentage": "agg_percentage_math"})
df_math=df_math.drop_duplicates
(subset='Q_5',keep=False)
df_math=df_math.set_index('Q_5')

df_questionnaire=pd.read_csv
('student questionnaire grade six.csv')
df_questionnaire=df_questionnaire
.drop_duplicates(subset='Q_0_7',keep=False)
df_questionnaire=df_questionnaire
.rename(index=str, columns={"Q_0_7": "Q_5"})
df_questionnaire=df_questionnaire
.set_index('Q_0_5')


df=df_questionnaire

df['agg_score_math']=df_math
['agg_score_math']
df['agg_percentage_math']=df_math
['agg_percentage_math']


#df['agg_percentage_Science']=df_science
['agg_percentage_Science']
#df['agg_percentage_Science']=df_science
```

```
['agg_percentage_Science']


### run this to get the number of common answers:
    df_clean=df.dropna()
# the ones that did math and questionnaire
    len(df[['agg_score_math']].dropna())

# the ones that did math, science and questionnaire
    len(df[['agg_score_math','agg_percentage_Science']]
    .dropna())

# the ones that did science and questionnaire
    len(df[['agg_percentage_Science']].dropna())

df.to_csv('merged_exam_grade.csv')
```

Regressions

Data: Johnson (2018); Software: R (R Core Team, 2019) In addition to the packages cited above, the 'stargazer' package (Hlavac, 2018) was used to help create Table **4.2** in LaTeXformat; and the 'tseries' package (Trapletti and Hornik, 2019) was used to run some of the tests to check the existence of the regression assumptions in the models .

```
setwd("/Users/******/Desktop/sudan")

final2 <- read.csv("/Users/*****/******/******
    /**********.csv",
                   na.strings=" ",header=TRUE)

library(tseries)

r1 <- lm(agg_percentage_Science ~ agg_percentage_math,
         data = final2)
```

```
r2 <- lm(agg_percentage_Science ~ agg_percentage_math
        +sai,
         data = final2)

r3 <- lm(agg_percentage_Science ~ agg_percentage_math
        +sai+school_type+gender,
         data = final2)

r35 <- lm(agg_percentage_Science ~ agg_percentage_math
        +sai+ school_type+ gender+father_work_q +mother_work_q,
         data = final2)

r4 <- lm(agg_percentage_Science ~ agg_percentage_math
        +sai+school_type+gender+father_work_q
         +mother_work_q+ father_ed+ mother_ed,
         data = final2)

AIC(r1,r2,r3,r35,r4, k=2)

par(mfrow=c(2,2))
plot(r4)

vif(r4)
lawstat::runs.test(r4$residuals)
acf(r4$residuals)
mean(sudan_model2$residuals)

library(stargazer)

stargazer( r1, r2, r3,r35, r4,
           font.size = "small",
           align = TRUE,
           column.sep.width = "-15pt")
```

Figure 5.1

Data: Johnson (2018); Software: R (R Core Team, 2019)

```
    ggplot(data=final2,
        aes(x=reorder
            (mother_work_type,agg_percentage_Science),
            y= agg_percentage_Science,
            ,fill=mother_work_type))
        + geom_boxplot()
        +theme_classic()
        +labs(y = "Science Achievement",
              x="Mother's Work Type",
              fill= "Work Types")
        + theme(text =
                element_text(size=12,
                             family="Garamond Bold"))
```

2018 NLA Achievement by State; Additional Correlation Plots in Appendix A

Achievement Data: Johnson (2018); Spending data: Siddig and Hassan (2016) ; Software : R R Core Team (2019)

```
#########################################################
##SCIENCE PLOTS (ACHIEVEMENT+ SPENDING CORRELATIONS)##
#########################################################

pdf("newscience_plots.pdf", width = par('din')[1],
    height = par('din')[2], onefile=T)


p0012 <- ggplot(data = sudan_state,
         aes (x=reorder(state,science),
            y=science, fill=science))

p0012   +geom_col(position="dodge2")
```

```
        + guides(fill=FALSE)
        + coord_flip()
        + scale_y_continuous(
                    labels = scales::percent_format
                                (accuracy = 1L),
                                limits=c(0,0.4))
        + labs(x= "State",
               y= "Science Achievement ",
               fill= "science")
        + geom_text(data = sudan_state,
                    aes(x=reorder(state,science),
                    y=science,
                    label = sprintf("%.02f %%",
                                science*100)),
                    position = position_dodge(width = 1)
                                        , hjust=-0.3)

        + theme_tufte()
        + theme(text =  element_text(size=12,
                        family="Georgia Bold"))
        + scale_color_hc()


p4 <- ggplot(data = sudan_state, mapping =
        aes (x=science,y=goods_services))


p4+ geom_point(mapping= aes(color=state))
    + geom_smooth(method="lm")
    +  theme_bw()
    + labs(y ="Spending on G&S",
            x="Science Achievement",
            color= "States",
            fill="States")
    +scale_x_continuous(labels =
            scales::percent_format
```

```
                (accuracy = 1L))
    +stat_cor(label.x = 0.16,
              label.y = 0.24)
    + stat_regline_equation(
                label.x = 0.16,
                label.y = 0.26) ##use this

p5 <- ggplot(data = sudan_state,
      mapping = aes (x=science,y=teachers))

p5  + geom_point(mapping= aes(color=state))
    + geom_smooth(method="lm")
    + theme_bw()
    + labs(y ="Spending on Teachers",
            x="Science Achievement",
            color= "States",
            fill="States")
    +scale_x_continuous(
            labels =
            scales::percent_format
                    (accuracy = 1L))
    +stat_cor(label.x = 0.16,
              label.y = 0.8)
    +stat_regline_equation(
              label.x = 0.16,
              label.y = 0.78) ##usethis

dev.off()
```

```
#########################################################
##MATH PLOTS (ACHIEVEMENT+ SPENDING CORRELATIONS)##
#########################################################

pdf("newmath_plots.pdf", width = par('din')[1],
    height = par('din')[2], onefile=T)


p1903 <- ggplot(data = sudan_state,
         aes (x=reorder(state,math),
            y=math, fill=math))

p1903   +geom_col(position="dodge2")
        + guides(fill=FALSE)
        + coord_flip()
        + scale_y_continuous(
                    labels = scales::percent_format
                                    (accuracy = 1L),
                                    limits=c(0,0.4))
        + labs(x= "State",
               y= "Math Achievement ",
               fill= "Math")
        + geom_text(data = sudan_state,
                    aes(x=reorder(state,math),
                    y=math,
                    label = sprintf("%.02f %%",
                                    math*100)),
                    position = position_dodge(width = 1)
                                              , hjust=-0.3)

        + theme_tufte()
        + theme(text =  element_text(size=12,
                        family="Georgia Bold"))
        + scale_color_hc()

p042 <- ggplot(data = sudan_state,
```

```
           mapping = aes (x=goods_services,y=math))

p042    + geom_point(mapping=
                aes(color=state,fill=state))
        + geom_smooth(method="lm")
        + theme_bw()
        + labs(x ="Spending on G&S",
               y ="Math Achievement",
               color= "States",
               fill="States")
        + scale_x_continuous(labels
                = scales::percent_format
                          (accuracy = 1L))
        + stat_cor(label.y = 0.6,
                   label.x = 0.01)
        + stat_regline_equation(label.y = 0.58,
                                label.x = 0.018)  ##use this

p051 <- ggplot(data = sudan_state,
        mapping = aes (x=math,y=teachers))

p051    + geom_point(mapping= aes(color=state))
        + geom_smooth(method="lm")
        +theme_bw()
        +labs(y ="Spending on Teachers",
              x ="Math Achievement",
              color= "States",
              fill="States")
    +scale_x_continuous(labels
                = scales::percent_format
                    (accuracy = 1L))
    + stat_cor(label.x = 0.3,
               label.y = 0.8)
    + stat_regline_equation(label.x = 0.3,
                            label.y = 0.78) ##usethis
dev.off()
```