HUMAN CAPITAL IN THE CONTEXT OF HIGH LEVELS OF INEQUALITY IN SOUTH AFRICA

Tamaryn Jean Friderichs

Student Number: G01R0006

Thesis presented for the degree of

DOCTOR OF PHILOSOPHY

in the

Department of Economics and Economic History

Faculty of Commerce

Rhodes University

June, 2021

Supervisors

Assoc. Prof. Gavin Keeton

Assoc. Prof. Mike Rogan

DEDICATION

To

My late father,

who asked for so little and gave so much.

ACKNOWLEDGEMENTS

Special thanks are due to my supervisors, Associate Professor Gavin Keeton and Associate Professor Michael Rogan, who provided valuable guidance in the writing of this thesis. Their support greatly contributed to what has been a fulfilling research experience. Furthermore, special mention is made of Dr Ferdi Botha who provided guidance in the initial conceptualisation of the thesis and statistical analysis. Thanks also to Dr Fabio Correa and Dr Roxana Chiappa for assistance with the statistical analysis. I am lastingly grateful to Associate Professor Nadine Oosthuizen for her invaluable mentoring, friendship and inspiration. I thank all my colleagues in the Economics Department for their continual encouragement and support.

My family have been the most important enablers of my achievements so far. My father's strength of character and the way in which he lived his life will always be an inspiration as I take heed of the resurrection promise that we will again be together in joy and without sorrow. I am ever grateful to my mother for her support in many ways throughout the years. I thank my husband and daughters for their daily inspiration, their patience, love and support; without their smiles and support this thesis would not have been completed. I am deeply grateful also to my parents-in-law for their unfailing support: it truly takes a village for a working mother to complete a doctoral thesis.

Finally, I thank God for His immense love, favour, grace and peace, in which I dwell every day.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVAITIONS

Adult Survival Rate	ASR
Changing Wealth of Nations	CWON
Coefficient of determination	CD
Comparative fit index	CFI
Data Reduction Technique	DRT
Demographic Household Survey	DHS
Generalised Redundancy Analysis	GRA
Gross Domestic Product	GDP
Harmonised test score	HTS
Household	HH
Household head	HHH
Household spouse	HHS
Human capital	HC
Human Capital Index	HCI
Human Capital Project	HCP
human capital theory	HCT
Human Development Index	HDI
Inclusive Wealth Report	IWR
Institute for Health Metrics and Evaluation	IHME
Instrumental variables	IV
Kaiser-Meyer-Olkin	KMO
Latent variable	LV
Multidimensional Poverty Index	MPI
National Income Dynamics Study	NIDS
National Satellite Accounts	NSA
Ordinary Least Square	OLS
Organisation for Economic Co-operation and Development	OECD
Partial Least Square	PLS

Principal Component Analysis	PCA
Progress in International Reading Literacy Study	PIRLS
Reduced Rank Regression	RRR
Root mean square error	RMSE
Root mean square error of approximation	RMSEA
Socioeconomic status	SES
South African Demographic Household Survey	SADHS
South African Labour and Development Research Unit	SALDRU
Structural Equation Modelling	SEM
Survey of Customer Finances	SCF
Sustainable Development Goal	SDG
Expected years of schooling	EYS
Luxembourg Income Study Database	LIS
Trends in International Mathematics and Science Study	TIMSS
Tucker-Lewis index	TFI
Under-5 mortality rate	U5MR
Under-5 stunting rate	U5SR
University of Cape Town	UCT
World Bank	WB
World Economic Forum’s Global Human Capital Index	WEF GHCI

CHAPTER 1

1 INTRODUCTION

1.1 Research context and scope of the study

Piketty's (2014) book titled "Capital in the Twenty-First Century" sparked widespread interest in global inequality, the distinction between wealth and income inequality and the economic, social and political processes accounting for changes in economic inequality over time. Piketty's (2014) study controversially stated that widening economic inequality is the normal state of affairs in capitalist societies. The return from capital/wealth (terms used interchangeably) will almost always outpace the returns from labour. In contrast with Piketty's (2014) thesis that the returns on non-human capital drive growing income inequality, economists such as Leibbrandt *et al.* (2012), Van der Berg (2014) and Hundenborn *et al.* (2016) have found that the labour market and human capital (HC) are the primary sources of income inequality. The research problem for this study stems from these contrasting views.

Piketty (2014) concluded that non-human capital drives income inequality by tracing historical changes in the concentration of income and wealth over 20 countries since the 18th Century. He identified that the rate of return on capital (r) is greater than economic growth (g) over the long-term. Thus wealth-to-income ratios (the share of capital income) rise, which gives rise to growing wealth inequality, causing persistent economic and social instability. He found that capital income is the primary source of income for the top 0.1% in the developed world and that wage levels are becoming increasingly skewed and not driven by productivity (Piketty, 2014). This view is echoed in the literature on financialisation, emphasizing the growing dominance of the financial sector and returns on capital in economies globally (Krippner, 2005, 2011). Piketty (2014) explained that periods where inequality has fallen, such as the middle decades of the 1900s, result from unusual occurrences such as the widespread destruction of capital during World Wars I and II.

The attention that Piketty's (2014) analysis received resulted in widespread support but also important criticism. Critics noted that Pikkety (2014) ignored HC from the measurement of wealth/capital, even though HC has repeatedly been recognised as a key driver of income and presumed higher productivity. HC is thus an essential element of household wealth (Becker, 1962; Castelló and Domenech, 2002; Arrow *et al.,* 2012; Piketty, 2014; Keeton, 2014; Blume and Durlauf, 2015 and Orthofer, 2017). Piketty's (2014) justification for the exclusion of HC from wealth and wealth inequality calculations is that HC cannot be traded on the market or owned by another person. Additional measurement difficulties of HC identified by other authors include that it is a multidimensional unobserved construct and the absence of a conclusive theoretical definition of HC result in a lack of a prescribed choice of observed variables (Di Bartolo, 1999; Costa, 2006; Lovaglio and Folloni, 2011 and Piketty, 2014)[1]. The measurement of HC is further complicated by the bidirectional relationship between HC and income/wealth.

Although HC is an input into income, as stated by the Mincerian wage function (Mincer, 1958 & 1974), as well as a component of household wealth (Becker, 1962), it is also an output of both wealth and income because the wealthy and those with higher incomes invest more in HC (Fernandez and Rogerson, 1992; Chiu, 1998 and Galor, 2011). Chani *et al.* (2012) thus suggest an unequal distribution of income/wealth may negatively impact HC inequality and vice versa. However, Blume and Durlauf (2015) stated that not including HC in the measure of capital/wealth simply because it is difficult to measure is not a sufficient reason for its exclusion. Wealth inequality calculations ignoring HC are 'meaningless'.

Summers (2014) in addition critiqued the policies promoted by Piketty (2014) to address inequality, which included a global wealth tax. Summers (2014) noted that Piketty accepts that such a global wealth tax is unlikely because of the incentive for individual countries to break ranks. Policies to reduce income inequality should instead, Summers (2014) suggest, aim to achieve bottom-up growth, for example, improved public investments and training of young people.

---

[1] The issues surrounding the definition of HC are further discussed on pages 13 and 14.

In the South African context, Arendse (2018) found that although a new wealth tax would increase the progressivity of the tax system, it would most likely not provide sustainable tax revenue to address economic inequality and would potentially harm economic growth. Furthermore, Leibbrandt *et al.* (2012) showed that in South Africa measured income inequality is not caused by income of a capital nature[2], but is generated in the labour market. Nonetheless, despite the criticisms of Piketty's (2014) economic analysis and policy recommendations, his findings provide a context for exploring the causes of inequality in South Africa.

South Africa is a country synonymous with very high levels of income[3] inequality with a gini coefficient ranging between 0.66 and 0.68 for the years 1993, 2008 and 2014 (Hundenborn *et al.,* 2018). With regards to distribution of wealth[4], Chatterjee *et al.* (2020) calculated South Africa to be the most unequal country in the world, with a gini coefficient of 0.95[5]. Unfortunately, inequality has not decreased over the 26 years since the end of Apartheid (Chatterjee *et al.*, 2020) but is rather increasing (StatsSA, 2019 and Spaull, 2020). Spaull (2020) aptly explained that "If SA was made up of 100 people and we lined them up from poorest to richest, the 10th person's income declined 156% and the 99th person's income (richest 1%) increased 48% from 2011 – 2015".

The unjust reality in South Africa is that life outcomes are determined largely by whether "a person is born a boy or a girl, black or white, in a township or leafy suburb, to an educated and well-off parent or otherwise" (World Bank, 2012:xii).

[2] Leibbrandt *et al.* (2012) analysed household income from five broad sources namely: labour market income (net of tax and including self-employment and wage earnings); income of a capital nature, income from government grants, remittance income and 'other' income.

[3] The National Income Dynamics Study (NIDS) derives household income as the sum of labour market income (net of taxes), government grants, other government income, investment income, remittances received, subsistence agriculture income and imputed rent for owner-occupied housing (Chinhema *et al.*, 2016). For the purpose of calculating inequality the household income is divided by the number of household residents to obtain a household per capita income, which is assigned to each household member (Leibbrandt *et al.*, 2012).

[4] Wealth (also known as 'net worth') is calculated as the difference between the market value of all assets and liabilities (Orthofer, 2017) and in NIDS household wealth is calculated by summing net financial wealth, net business equity, net real estate equity, value of vehicles, total value of pension/retirement annuities and livestock wealth (Chinhema *et al.*, 2016).

[5] HC was not included in the measure of wealth (Orthefor, 2014)

Spaull (2019) found that predicting a child's future socioeconomic outcomes (for example employment, income and education) in South Africa can be done with a certain level of accuracy by the age of seven. Many talented South African citizens are thus not likely to reach their full potential, negatively impacting much needed social and economic development (Keeton, 2014). These findings for South Africa are consistent with social reproduction theory which poses that the social stratification order is reinforced by the education system (Bourdieu and Passeron, 1977). Likewise, Fedderke *et al.* (2003), Liebbrandt *et al.* (2010) and van der Berg (2010) concluded that to reduce inequalities in South Africa, increases in HC, particularly for those presently poor, requires additional policy attention and further research.

The concept of HC is shrouded with controversy, with it previously having been considered morally wrong to value humans as resources contributing to national wealth (Schultz, 1959 and Folloni and Vittadini, 2010). The concept of HC in economics dates back to the seventeenth century, with Sir William Petty (1623-1687) being one of the first economists to define and measure it (Folloni and Vittadini, 2010). Alfred Marshall (1890/1922:564) highlighted the importance of HC to national wealth, stating that the "most valuable of all capital is that invested in human beings". On a disaggregated level, this would translate to the HC of each individual/household being, on average, its most valuable asset.

Before the twentieth century, research regarding HC was similarly centered around enhancing national wealth accumulation (Folloni and Vittadini, 2010). A shift in thought occurred mid-century, with the introduction at the University of Chicago of microeconomic theories that regarded HC as a rational investment choice (Schultz, 1959 & 1961, Becker, 1962 & 1964 and Mincer, 1958 & 1974). These authors were thereafter considered the 'founders' of HC theory.

The 'new' theories were concerned with the "activities that influence future real income through the embedding of resources in people; that is measuring HC as the investments in HC" (Becker, 1962:9). In line with this thinking, modern economic theories, including the New Growth Theory, New Institutionalist Economics and

Sen's Capabilities approach, highlight the need to move beyond relying on non-human capital for sustainable development and emphasise the requirement for additional investments in HC (Evans, 2010). However, accurately measuring and analyzing this HC remains a challenge, given the lack of a conclusive definition, the reality that it cannot be traded on the market or owned by another person (Piketty, 2014), as well as data limitations at a disaggregated and household level (D'Souza *et al.*, 2019).

More recent definitions of HC have been extended to incorporate non-market activities (Le *et al.*, 2005). Therefore, a broad definition of HC has been recognised as "the knowledge, skills, competencies and attributes embodied in individuals that facilitate the creation of personal, social and economic well-being" (OECD, 2001:18). This was an extension of the earlier 1998 definition by the Organisation for Economic Co-operation and Development (OECD) as "the knowledge, skill, competencies and attributes embodied in individuals that are relevant to economic activity" (OECD, 1998:9). The definition of what HC is and how it should be measured continues to be developed.

Oxley *et al.* (2008) identified three general approaches to measuring HC. They include an income-based measure (originally modeled by Farr (1853) using macro data and later revived by Weisbrod (1961) using cross-sectional micro data), a cost-based measure (Kendrick, 1976 and Eisner, 1985, 1989) and, lastly, a group of education-based (i.e. indicator-based) methods (Barro and Lee, 1993; Hanushek and Kimko, 2000 and Wößmann, 2003)[6]. The income-based method measures a population's present value of life-time earnings and thus estimates HC by focusing on future investment returns. The cost-based approach focuses on measuring the inputs used in producing HC, including the costs of raising and educating people. The education-based method proxies HC through indicators such as literacy rates, school enrolment rates and educational attainment.

[6] Since then an integrated latent-variable approach has been developed with Dagum and Slottje (2000) being one of the first examples.

Liu and Fraumeni (2021) categorised the HC measures into two broad approaches namely indicator-based and monetary measures. Towards providing researchers with comparable HC measures, Liu and Fraumeni (2021) discussed the methods and findings of six HC measures each of which are estimated across at least 130 countries. The monetary approaches discussed by Liu and Fraumeni (2021) are those developed to help government planning by including HC and nonhuman capital in a framework of national wealth accounting. The World Bank's report titled "Changing Wealth of Nations 2018: Building a Sustainable Future (CWON)" included a monetary value for HC for 141 countries over the period 1995 – 2014 by applying the Jorgenson-Fraumeni approach (Liu and Fraumeni, 2021). A second HC monetary measure is that presented in the biennial Inclusive Wealth Report (IWR), which included HC monetary values for 140 countries over the period 1990 – 2014. The HC measure was estimated by following a framework developed by Klenow and Rodriguez-Clare (1997) and Arrow *et al.* (2012, 2013).

Indicator-based measures discussed by Liu and Fraumeni (2021) include four largescale composite indexes for HC. The indexes include the World Bank's Human Capital Index (HCI), United Nation's Human Development Index (HDI)[7], the Institute for Health Metrics and Evaluation Human Capital Index (IHME HCI)[8] and the World Economic Forum's Global Human Capital Index[9] (WEF GHCI). The indexes use various macroeconomic education and health indicators as the two basic dimensions of HC towards estimating a counties' distance from optimal HC (the distance from maximum HC attainable). The index applied in this study is the HCI (World Bank, 2018a), developed by the World Bank's Human Capital Project (HCP) in an attempt to improve the analysis of HC. The HCI measures, at a country level, the HC "a child born today can expect to attain by her 18th birthday, given the risks of poor health and poor education where she lives" (World Bank, 2018a:3). The

[7] The HDI was developed to measure human development as a composite index and included in the first Human Development Report written by a team at the UNDP led by Mahbub ul Haq (UNDP, 1990). South Africa ranks 114th out of 189 countries (UNDP, 2020)
[8] The IHME HCI measures the "expected years lived from age 20 to 64 years and adjusted for educational attainment, learning, and functional health status" (Lim *et al.*, 2018: 1217). South Africa ranks 144th out of 195 countries.
[9] The WEF GHCI has been constructed since 2015 with the updated methodologies and most recent estimates presented for 130 economics for 2017 in The Global Human Capital Report 2017 (World Economic Forum, 2017). South Africa ranks 87th out of 130 countries.

purpose of the index was to provide policymakers with an understanding of how much development and growth is being lost given low levels and an unequal distribution of HC (World Bank, 2019a). As such, the HCP aimed to work with countries to ensure better investments into HC. South Africa has a low ranking HCI estimate of 126 out of 157 countries (World Bank, 2018a), confirming the importance of further research on understanding the multidimensional nature of HC in South Africa at a microeconomic level. The socioeconomic status (SES) HCI has since also been developed by the HCP with the aim of measuring the distribution of HC within a country at different disaggregated levels.

Because education is widely considered to be a fundamental dimension of HC, and also because it is easily measured through available data, education attainment has formed the basis of most HC research in South Africa (Lam, 1999; Liebbrandt *et al.*, 2010; Van der Berg, 2002 & 2010; Pellicer and Ranchhod, 2012; Branson and Leibbrandt, 2013 and Wanka, 2014). However, education is simply a proxy for HC, and the accuracy and continual relevance thereof has been questioned (Le *et al.*, 2005 and Laverede-Rojas *et al.*, 2019).

Since the late twentieth century, an additional approach to measuring HC has evolved namely a latent-variable approach. The approach integrated the indicator- and income-based approaches by extending the measured latent HC into monetary terms. Dagum and Slottje (2000) proxied household HC as a "multidimensional non-observable construct generated by personal ability, home and social environments, and investments in the education of the household head and spouse whose effects are indirectly measurable by means of the present value of a flow of earned income throughout an individual's life span whose effects are reflected in household income" (Dagum *et al.*, 2007:581). In a literature review titled "Human Capital Measurement: A survey" Folloni and Vittadini (2010) concluded that HC should be measured as a latent construct[10] in a Structural Equation Modelling (SEM) framework (Laverede-Rojas *et al.*, 2019).

[10] The late 20th and 21st century research on the topic of measuring and analyzing HC as a latent variable and its distribution has mainly being contributed by Dagum (1994), Dagum and Vittadini (1996), Di Bartolo (1999), Dagum and Slottje (2000), Vittadini *et al.* (2003), Foldvari and van Leeuwen (2005), Dagum *et al.*, 2007, Vittadini and Lovaglio, 2007, Lovaglio (2008), Lovaglio (2010), Lovaglio

Given the complex nature of HC and a lack of a theoretically explicit definition (Costa, 2006), all methods have identified shortfalls and continue to be refined. This search for improvement has been expressed in literature reviews, including those by Le *et al.* (2005), Oxley *et al.* (2008), Folloni and Vittadini (2010), Slottje (2010) and Liu and Fraumeni (2021).

Following on from this review of the criticisms of Piketty's (2014) study, the challenges of measuring HC and the importance of analysing HC in the South African context of very high levels of inequality, the research gap and subsequent goals of this research are identified.

1.2 Research goals

The main research goal is: To measure human capital at a disaggregated level and determine the importance of human capital in explaining income inequality in South Africa.

The research sub-goals are:

1) To construct and compare a set of measures of HC at a disaggregated level in the South African context;
2) To measure the distribution of HC according to socioeconomic status in South Africa;
3) To determine the extent to which human capital explains income inequality in South Africa.

1.3 Justification/Rationale for the Study

This thesis contributes to the critical literature on HC, the distribution of HC by socioeconomic status and how unequal distribution of HC feeds into income

and Vittadini (2011) and Fattore *et al.* (2017). Despite data constraints and method challenges (Di Bartolo, 1999) the researchers continue to develop the methods striving towards obtaining the most accurate measure of HC.

inequality. HC is the embodiment of the productive resources facilitating the creation of social and economic well-being. The measurement of this resource is however challenging given the multidimensional unobserved nature of HC and its bidirectional relationship with income. Different approaches are available to measure HC at a disaggregated level. Each approach has alternative definitions of HC and thus has different data requirements for its measurement.

This study uses three measures, namely the latent-variable, education-based and SES-HCI approaches, to understand the distribution of HC in South Africa by socioeconomic status in South Africa. The study results provide valuable insights into the limitations, contributions and estimated distribution of HC by socioeconomic status for each measurement approach and how HC feeds into income inequality in South Africa. The study, therefore, adds to the body of knowledge by reconciling the meaning and measurement of HC, estimating HC using three different measurement approaches and the resultant distribution of HC by socioeconomic status, and analysing the extent to which HC explains income inequality in South Africa.

1.4 Data

Primary data from the National Income Dynamics Study (NIDS) dataset is employed. The NIDS is a panel dataset compiled biennially by the South African Labour and Development Research Unit (SALDRU) based at the University of Cape Town (UCT). It is the result of an initiative by the South African Presidency to analyse changes in the wellbeing of South Africans over time. The first wave of the dataset, released in 2008 (NIDS, 2018), was administered to a nationally representative sample[11] of approximately 28 000 individuals in 7 296 households but grew with each wave given children born and adopted, to a total of 37 000 individuals and 13 719 households included in the most recent dataset. Wave 5, was conducted in 2017 and released in 2018 (Brophy *et al.*, 2018) and is employed as a cross-sectional dataset in each of the studies in the current research. In addition, Wave 1 is

[11] The total number of individuals successfully interviewed in Wave 1 and Wave 5 were 26 776 and 37 368 respectively (Brophy *et al.*, 2018).

employed in Chapter 5, to identify changes over the nine years from 2008 to 2017. Different samples were selected from the NIDS dataset for each of the chapters depending on data requirements of the different measurement methods employed to measure HC. The sample data selected is explained in each chapter.

The main reason for employing the NIDS dataset was the comprehensive nature of the socioeconomic data collected[12], providing a combination of labour and other socioeconomic indicators data. The data was also applicable to each of the measurement approaches employed in the study, allowing for comparative findings[13].

1.5 Structure of the remaining chapters

The research goals were achieved by completing four academic articles, which have been submitted to accredited academic journals. The four articles are each presented as a separate chapter (Chapters 2, 3, 4, 5), which work together to achieve the research sub-goals. Chapter 6 reconciles and synthesizes each article's findings to provide concluding comments and address the main research goal.

To develop a framework for constructing and comparing a set of measures of HC at a disaggregated level in South Africa, the first article presented in Chapter 2[14] is a literature review article titled "Reconciling the meaning and measurement of HC as a determinant of socioeconomic inequalities". The article sets out the suggested definitions of HC and aims to reconcile the meaning and measurement of HC at a disaggregated level. This is necessary for any analysis of the distribution of HC. It is noted that, because education is a fundamental dimension of HC, and because of the easy accessibility of education attainment statistics (Barro and Lee 2013), years of education are most commonly used as the sole indicator of HC (Lee and Lee, 2018). However, HC's unobserved and multidimensional nature has led some

[12]Examples of variables which will be used from the NIDS dataset include school administration data, grade repetition, school quintile, literacy, educational attainment, employment history, income, parents' education, stunting, mortality history, assets, debt and health.
[13] For a more detailed description of the data and survey methodology and weights, the interested reader is referred to www.nids.uct.ac.za.
[14] The article is currently under consideration in the Journal of Economic Inequality.

analysts (most notably Dagum and Slottje, 2000) to reject this reliance on educational attainment. Dagum and Slottje (2000) thus insist, instead, that HC should be measured as a latent variable, within a SEM framework, integrating the indicator- and income-based approaches to measure a monetary value for household HC (Dagum and Slottje, 2000).

Chapter 2 therefore presents the aims, method, data and variables employed in the previous studies of HC as a latent variable in a SEM framework. The methodological strengths and weaknesses of these studies are examined. The findings are presented in a table used to develop the model used in Chapter 3 and to aid future researchers in model development. The focus is on 21st century research studies by authors including Dagum and Slottje (2000); Dagum *et al.* (2003); Vittadini *et al.* (2003); Foldvari and van Leeuwen (2005); Dagum *et al.* (2007); Vittadini and Lovaglio (2007); Lovaglio (2008); Lovaglio (2010); Folloni and Vittadini (2010); Slottje (2010), Lovaglio and Vittadini (2014); Lovaglio *et al.* (2016); and Fattore *et al.* (2018). The study builds on the previous literature review articles on measuring HC by Le *et al.* (2005); Oxley *et al.* (2008); Folloni and Vittadini (2010) and Slottje (2010). Furthermore, supplementing HC theory with the human capabilities approach is explored towards conceptualising HC.

To achieve the second research goal of measuring HC distribution across individuals, households and subgroups in South Africa, Chapters 3 and 4 present two research articles in which HC and the distribution of HC by socioeconomic status indicators are measured. In Chapter 3 they are estimated using the latent-variable and education-based approaches. The SES-HCI approach is used in Chapter 4. Chapter 3 presents an article titled "Measuring human capital distribution in South Africa using a latent-variable approach".[15] It firstly estimates the latent construct HC using a confirmatory factor analysis measurement approach. The observed indicators used in the measurement of latent HC are educational attainment, parental education attainment, school quality (measured

[15] The article is under review in the South African Journal of Economics. The co-author is Dr Fabio Correa.

using school quintile[16]) and health. These indicators were selected based on social reproduction and microeconomic HC theory, as well as the evidence of previous empirical research. Secondly, the article compares the profile of HC by socioeconomic status, namely income, gender, race and geographical area, using the latent-variable and education-based approaches (using education attainment as a proxy) to estimate HC. Given the intergenerational transmission of life outcomes in South Africa (Spaull, 2019), it is expected that parental education is a significant contributing factor to the HC latent construct.

Chapter 4 estimates HC in South Africa using the World Bank's HCI disaggregated by socioeconomic status (SES-HCI) indicators, namely, income, school quintile, geographical area, gender and race. The SES-HCI was developed by the World Bank (2018a) to analyse inequalities within countries and is applied in this article to identify the gaps in HC development in South Africa. The article is titled "Measuring human capital in South Africa using a socioeconomic status human capital index approach"[17]. The article aims to provide evidence-based suggestions towards addressing HC gaps and highlight the limitations of applying the SES-HCI to study inequality in South Africa. Chapters 3 and 4 thus provide comparative estimates of the distribution of HC by socioeconomic status in South Africa measured using three alternative approaches. Both studies employ the NIDS wave 5 dataset.

Chapter 5 addresses the final sub-goal of determining the extent to which HC explains income inequality in South Africa. HC has increasingly been identified in the literature as a driver of economic development with the potential to reduce income inequality (Evans, 2010 and Lee and Lee, 2018). In South Africa, income inequality has been identified as being driven by the labour market in which HC is employed (Leibbrandt *et al.,* 2012). No significant decreases have occurred in the

[16] Public schools in South Africa are classified from quintile 1–5 (Q1-Q5) with quintile1 being the poorest and quintile 5 being the wealthiest. No-fee paying schools are categorised as Q1-Q3, low-fee paying schools as Q4 and fee-paying schools as Q5 schools, which mainly consist of the formerly White 'Model C' schools (Spaull, 2019).

[17] The article has been published online through the international journal Development Southern Africa. The co-authors are Prof Keeton and Prof Rogan. Available at: https://doi.org/10.1080/0376835X.2021.1941779

level of inequality over the 26 years since the end of apartheid (Chatterjee *et al.*, 2020). Fields' regression-based decomposition method is employed in this study, which allows for identifying key HC and socioeconomic drivers of income inequality. The article is titled "Decomposing household income inequality in South Africa by human capital determinants"[18].

The Fields method provides an additional approach to identify whether using educational attainment as a proxy for HC is justified when analysing income inequality in South Africa. The HC indicators included in the study are education attainment, self-perceived health and schooling quality. The analysis is conducted at the household level, based upon the assumption that resources are shared amongst household members (Pieters, 2011). Proxies for household education attainment are explored. Changes over time are presented with the analysis spanning a nine-year period from 2008 to 2017, drawing from the NIDS wave 1 and wave 5 datasets. Furthermore, the study is conducted for the total South African household population and subgroups, namely gender, race and geographical area, each of which plays a major role in labour market outcomes and socioeconomic well-being.

Chapter 6 summarises what has been learned through the research papers presented towards a better understanding of the measurement and distribution of HC in South Africa by socioeconomic status and its role in explaining inequality.

[18] The article is currently under review in the South African Journal of Economics. The co-authors are Prof Keeton and Prof Rogan.

CHAPTER 2

2 RECONCILING THE MEANING AND MEASUREMENT OF HUMAN CAPITAL AS A DETERMINANT OF SOCIOECONOMIC INEQUALITIES[19]

Abstract[20]

The definition of what human capital (HC) is and how it should best be measured continues to be refined. Because education is a fundamental dimension of HC and the ready availability of education attainment statistics, years of education are most commonly used as the sole indicator of HC. The complex nature of HC led Dagum and Slottje (2000) to insist that HC should be measured as a latent variable. This study reviews previous studies using the latent-variable approach to specify the considerations required when developing a multidimensional conceptual model reconciled with the meaning of HC. The nexus between the latent-variable approach to measuring HC and the human capabilities approach is explored towards conceptualising HC. It is noted that the theoretical framework, local research context and data availability are vital in ensuring meaningful measurement of HC and analysis of socioeconomic inequalities.

Keywords: human capital; education; latent variable; human capabilities

[19] The article is under consideration the Journal of Economic Inequality.

2.1 Introduction

Modern economic theories, including the New Growth Theory, New Institutionalist Economists and Sen's Capabilities approach, highlight the need to move beyond non-human capital for sustainable development. Instead, additional investments in human capital (HC) are required (Evans, 2010). However, given the lack of a conclusive definition, the reality that it cannot be traded on a market or owned by another person (Piketty, 2014), as well as data limitations at a disaggregated and household level (D'Souza *et al.*, 2019) means that accurately measuring and analysing HC remains a challenge. According to Folloni and Vittadini (2010), thinking in economics about HC dates back as far as the seventeenth century, when William Petty (1623-1687) first tried to define the concept and then measure it. Folloni and Vittadini (2010) note that research regarding HC before the twentieth century was largely centred around enhancing national wealth accumulation. In line with this goal, Petty (1690) argued that factors other than land and property should be included when determining the wealth of a nation. Instead, he believed that labour was the 'father of wealth' of a nation (Folloni and Vittadini, 2010). The motive for Petty's (1690) attempt to measure HC was to demonstrate how powerful England was. He calculated the addition to national wealth from its HC to be valued at 520 million pounds, or eight pounds per capita (Le *et al.*, 2005).

Initially, the discussions regarding HC were shrouded in controversy as it was considered morally wrong to value humans as mere resources contributing to national wealth (Schultz, 1959; Folloni and Vittadini, 2010). A shift in thought occurred with the introduction of microeconomic theories that regarded investment in HC as a rational choice (Schultz, 1959 & 1961; Becker, 1962 & 1964; Mincer, 1958 & 1974). These authors were thereafter considered the founders of modern human capital theory (HCT). Their 'new' theories were concerned with the "activities that influence future real income through the embedding of resources in people", that is, "investing in human capital" (Becker, 1962:9). The Mincerian model, which bypassed quantitatively estimating HC (Dagum and Slottje, 2000), focused on how the investments in HC noted by Becker (1962) to include factors such as education, on-job training and healthcare impacted earnings. However, the

Mincerian function was criticised because of its omission of important variables that determine wages (Lovaglio, 2010), such as the quality of education received. The Mincerian function has thus been extended and "the term 'earnings function' has come to mean any regression of individual wage rates or earnings on a vector of personal, market, and environmental variables thought to influence the wage" (Willis, 1986 as cited in Dagum and Slottje, 2000:71).

Sen (1997) however noted the earnings function and more broadly HCT has a restricted focus when studying human capabilities, given HCT is concerned with the indirect value of humans as 'capital' in the production process. Sen (1997) highlighted the distinction between human capabilities and HC, identifying that although related areas placing humans at the centre of attention of development and concerned with the abilities of people, human capabilities are more inclusive, valuing both indirect and direct values of human abilities. The direct values include characteristics such as a person's abilities and social and economic status related to the ability to directly achieve the freedom to live a life he/she values. The human capabilities approach thus focuses on "the ability of human beings to lead lives they have reason to value and to enhance the substantive choices they have" (Sen, 1997: 1959), and thus achieve functionings valued. Sen (1997) therefore aimed to evaluate social arrangements based on capabilities which combines the two concepts functionings (the achievement of beings and doings valued) and freedom of choice. Sen (1997) recognised that a broader definition of HC might cover both direct and indirect values. For example, education attainment may be valued as the achievable functioning of making informed life choices (direct value) and the functioning of being able to provide a service that increases economic productivity and individual income (indirect value). Considering the value of education, Chiappero-Martinetti and Sabadash (2014) thus identified the narrow focus of the HCT as the potential entry point for the capabilities approach, concluding that, although subject to methodological challenges, the HCT and capabilities approach paradigms have the potential to complement each other.

The definition of what HC is and how it should be measured is inconclusive and continues to be refined and thus a measurement challenge (Costa, 2006 and Piketty, 2014). Schultz (1961) defined HC as: "…all human abilities to be either innate or acquired" and suggested that any "(a)ttributes … which are valuable and can be augmented by appropriate investment will be human capital" (Schultz, 1961:21). Becker (1964) expanded on Schultz's (1961) definition stating that HC is the "knowledge, information, ideas, skills, and health of individuals". Becker's (1964) definition thus added the extra dimension of health. This expansion was significant in that it recognised the multidimensional nature of HC and the context in which it is developed are now recognised as important in contemporary research (CIPD, 2017). Thus, whereas the earlier definitions of HC focused on how investment in HC could improve productivity in terms of market related earnings as applied in the Mincerian wage function (Le *et al.*, 2005), more recent definitions have been extended to incorporate non-market activities. Therefore, a broader definition of HC today is "the knowledge, skills, competencies and attributes embodied in individuals that facilitate the creation of personal, social and economic wellbeing" (OECD, 2001:18) as opposed to the earlier 1998 definition by the OECD as "the knowledge, skill, competencies and attributes embodied in individuals that are relevant to economic activity" (OECD, 1998:9)[21]. The expanded OECD (2001) definition of HC is therefore more aligned with Sen's human capabilities approach, given value is assigned to outcomes beyond commodity production.

Economic growth, although important, is thus not the end, but the means to live a freer and more fulfilling life. Incorporating social and personal wellbeing into the definition of HC, supplements the meaning of HC by recognising that human qualities valued are not restricted to those employed in production but also social and personal wellbeing, such as health, respect, and knowledge etc[22]. Analysing HC requires the measurement approach be reconciled with the definition of HC, specified as per the theoretical framework employed, and results interpreted within these specifications.

---

[21] As identified by Adam Smith, HC has a role to play in quality of life over and above economic well-being, with a distinct contribution to social well-being (Stiglitz *et al.*, 2009).

[22] Sen (1997) noted that development requires improved opportunities which expand human capabilities towards living freer more worthwhile lives.

Oxley *et al.* (2008) identify three general approaches[23] to measuring HC namely the income-, cost- and indicator-based[24]. The income-based approach is considered prospective and values HC as the individual's lifetime earnings generated in the labour market. This is the original simplistic approach used by Petty (1690) and further developed into a scientific model by Farr (1853) (Le *et al.,* 2005). However, valuing HC using income alone neglects to account for the quality of life, which is distinct from the economic outputs of HC. Stiglitz *et al.* (2009) explained the importance of quality of life, by referring to Adam Smith, noting a lack of education of a boy does not only reduce future earnings but "when he is grown up he will have no ideas with which he can amuse himself" (Stiglitz *et al.*, 2009:165).

In contrast to the income-based approach, the cost-based approach is retrospective, assuming the value of HC of individuals to be the cost (expenditure on) of accumulating 'wealth' in terms of HC (Folloni and Vittadini, 2010). The limitations of the cost-based approach highlighted by Chiappero-Martinetti and Sabadash (2014) include: the assumed direct relationship between investment and the quality of output; capital is valued by the costs incurred rather than by the demand; and it does not account for non-market inputs for example family contributions. Despite the shortfalls, the income- and cost-based approach to measuring HC have been used when studying HC in National Satellite Accounts (NSA)[25] (Gu and McDonald, 2016), with education indicators suggested as a benchmark or complementary statistics (United Nations, 2016). The NSA has been introduced to extend national accounting analysis beyond traditional economic measures such as Gross Domestic Product (GDP) by accounting for social assets (including HC) towards assessing development. The challenges of incorporating HC in NSA included the measurement

[23] Income-based measures were originally modeled by Farr (1853) using macro data and later revived by Weisbrod (1961) using cross-sectional micro data; the cost-based measure has been developed by authors including Kendrick (1976) and Eisner (1985, 1989); and, the education-based (i.e. indicator-based) method has been expanded by authors including Barro and Lee (1993), Hanushek and Kimko (2000) and Wößmann (2003).
[24] Oxley *et al.* (2008) specifically discusses the education-based approach which falls under the broader indicator-based approach (United Nations, 2016).
[25] In 2008, French President Sarkozy, commissioned a task team, Commission of the Measurement of Economic Performance and Social Progress, to address the unsatisfactory state of national statistical information about the economy and society (Stiglitz *et al.,* 2009).

difficulties of HC and the inability of a conceptual framework of HC in one country to capture the reality of another country.

In the context of the indicator-based approach, Oxley *et al.* (2008) discusses the education-based approach to measuring HC. The proxies employed in the education-based approach include educational attainment or school enrolment and has been extended to including measures of educational quality (Wößmann, 2003). The Barro and Lee Educational Attainment Dataset contributed significantly to the education-based method, measuring HC on a macro level by providing data for 146 countries in 5 years intervals, disaggregated by gender, over the time period 1950 – 2010 (Barro and Lee, 1996). Researchers who have used the dataset to analyse the relationship between HC and income inequality have included Castelló and Domenech (2002 & 2017), Chani *et al.* (2014), Shahpari and Dovoudi (2014), Johansen (2014) and Lee and Lee (2018). Education attainment was used as the simplistic proxy for HC in the studies despite the definition of HC requiring the inclusion of additional elements. Education attainment as a measure of HC has been criticised given it does not account for the quality of skills and knowledge and other sources of learning such as on-the-job training (Wößmann, 2003). Thus, restricting human abilities to education attainment results in additional information about the accumulated HC not being captured (Wigley and Akkoyunlu-Wigley, 2005 and Chiappero-Martinetti and Sabadash, 2014). Furthermore, the education attainment does not capture human diversity and difference in opportunities presented as labour market discrimination and occupational segregation, for example, career and wage gaps. Thus, two individuals with the same education attainment may experience different job opportunities and the ability to improve their socioeconomic status resulting in persistent inequalities (Chiappero-Martinetti and Sabadash, 2014).

Liu and Fraumeni (2021) categorised the HC measures into two broad approaches namely indicator-based and monetary measures. Towards providing researchers with comparable HC measures, Liu and Fraumeni (2021) discussed the methods and findings of six HC measures each of which are estimated across at least 130 countries. The monetary approaches discussed by Liu and Fraumeni (2021) are those developed to help government planning by including HC and nonhuman capital in a framework of national wealth accounting. The World Bank's report titled "Changing Wealth of Nations 2018: Building a Sustainable Future (CWON)" included a monetary value for HC for 141 countries over the period 1995 – 2014 by applying the Jorgenson-Fraumeni approach (Liu and Fraumeni, 2021). A second HC monetary measure is that presented in the biennial Inclusive Wealth Report (IWR) which included HC monetary values for 140 countries over the period 1990 – 2014. The HC measure was estimated by following a framework developed by Klenow and Rodriguez-Clare (1997) and Arrow *et al.* (2012, 2013).

Indicator-based measures discussed by Liu and Fraumeni (2021) include four largescale composite indexes for HC. The indexes include the World Bank's HCI[26], United Nation's Human Development Index (HDI)[27], the Institute for Health Metrics and Evaluation Human Capital Index (IHME HCI)[28] and the World Economic Forum's Global Human Capital Index[29] (WEF GHCI). The indexes use various macroeconomic education and health indicators as the two basic dimensions of HC towards estimating a counties' distance from optimal HC (the distance from maximum HC attainable).

The unobserved complex nature of HC led Dagum and Slottje (1994 & 2000) to state that HC is statistically a latent variable and thus should be measured as such. HC, they suggest, is statistically defined as a latent variable as it is a multidimensional non-observed construct formed by indicators that include "personal ability, social

---

[26] The World Bank's HCI was developed by the Human Capital Project (HCP), to measure the HC "a child born today can expect to attain by her 18th birthday, given the risks of poor health and poor education where she lives" (World Bank, 2018a:3). Survival (under-5 mortality rates), education (learning adjusted years of education) and health (under-5 stunting and adult survival) components are used to measure the HCI.

[27] The HDI was developed to measure human development as a composite index and included in the first Human Development Report written by a team at the UNDP led by Mahbub ul Haq (UNDP, 1990). The HDI embodied the capabilities approach (Sen, 1985 and Nussbaum 1988) and human development theoretical concepts. The HDI measures critical elements of human development, estimating three dimensions: a long and health life, knowledge, and a decent standard of living by means of dimension indices namely life expectance index, knowledge index, and gross net income (GNI index).

[28] The IHME HCI measures the ""expected years lived from age 20 to 64 years and adjusted for educational attainment, learning, and functional health status" (Lim *et al.*, 2018: 1217).

[29] The WEF GHCI has been constructed since 2015 with the updated methodologies and most recent estimates presented for 130 economics for 2017 in The Global Human Capital Report 2017 (World Economic Forum, 2017)

and home environments, investments in the education of the household head and spouse" (Folloni and Vittadini, 2010:267), whose importance are reflected in household (HH) income (Vittadini *et al.,* 2003). The approach integrated the indicator- and income-based approaches. In the 21st century numerous studies[30] have been conducted on measuring and analyzing HC as a latent variable and its distribution. The researchers have sought to develop the methods for obtaining the most accurate measure of HH HC and the distribution thereof (Slottje, 2010) to supplement public policy analysis. For a discussion on the history of HC and previous literature reviews on the measurement approaches of HC as a latent variable, see Le *et al.* (2005), Oxley *et al.* (2008), Folloni and Vittadini (2010) and Slottje (2010), which this research aims to build on.

Constructing a comprehensive measure for HC is vital for researching a range of economic questions, particularly given the recognised importance of the growing knowledge economies (Le *et al.*, 2003 & 2005 and Oxley *et al.,* 2008). Addressing socioeconomic inequalities would benefit from an accurate measure of HC given the bidirectional positive relationship between income/wealth and HC; unequal distribution of income/wealth negatively impacts HC inequality, and *vice versa* (Chani *et al.,* 2014). Amongst others, Checchi (2001) and De Gregorio and Lee (2002) studied the relationship between HC inequality and its impact on income inequality, using education attainment as a proxy for HC, despite the acknowledged weaknesses of this approach (Castelló and Doménech, 2002 & 2017). In addition, HC is often measured at an aggregate level (e.g. the monetary and index measurements discussed by Liu and Fraumeni (2021)), partly because of a lack of comprehensive HH level data. However, in the context of high levels of income inequality and poverty, measurement at a disaggregated (D'Souza *et al.,* 2019) individual/HH level is critical (Pieters 2011).

Furthermore, the HH level is the basis for government planning and service delivery (Pieters, 2011). Resources (including HC) are, at least to some extent, shared among

---

[30] The studies are by authors including Dagum and Slottje (2000); Dagum *et al.*, 2003); Vittadini *et al.* (2003); Foldvari and van Leeuwen (2005); Dagum *et al.* (2007); Vittadini and Lovaglio (2007); Lovaglio (2008); Lovaglio (2010); Folloni and Vittadini (2010); Slottje (2010), Lovaglio and Vittadini (2014); Lovaglio *et al.* (2016); and Fattore *et al.* (2018).

HH members and thus, individual capabilities also determined by HH welfare. As a result, HH per capita measures are a more accurate[31] calculation of inequality than measures on an individual level and reflect an important characteristic of the development process (Pieters 2011). Consequently, the latent-variable approach, which measures HC at a disaggregated level, should be explored when investigating public investments and socioeconomic inequalities. The Luxembourg Income Study Database (LIS) was established to improve micro-level data availability, making available "household- and person-level data…on labour income, capital income, pensions, public social benefits (excl. pensions) and private transfers, as well as taxes and contributions, demography, employment, and expenditures" (LIS Database, 2019).

However, Costa (2006) highlights that for latent variable estimates of HC, a lack of an explicit theoretical definition of HC means an absence of a prescribed choice of observed variables. Thus, although improvements in data availability and accuracy have allowed for extended estimates of HC, when measuring HC as an unobserved construct, the observed variables and therefore estimates of HC are not clearly defined.

Aligning HCT and the capabilities approach towards supplementing the analysis of HC would furthermore require a reconsideration of the meaning and variables. Incorporating variables according to the wellbeing process[32], which accounts for the value of human abilities beyond economic benefits, human diversity, and differences in opportunities, should be considered (Chiappero-Martinetti and Sabadash, 2014). However, it is expected that data limitations and given, as stated by Chiappero-Martinetti and Sabadash (2014), personal economic and social wellbeing are resistant to measurement, estimating HC conceptualised to account for both the indirect and direct values of human abilities at a disaggregated level would be

[31] When measuring HC at the HH level it is however important to take note of the warning, by Rogan (2016), against using headship or a single reference person to classify HHs. Thus, to measure HH HC it is important that the definition of HH HC and the proxies used capture the meaning of HH HC.
[32] Chiappero-Martinetti and Sabadash (2014) highlighted the wellbeing process which has become quite common the capabilities literature which recognises the mechanism through which endowments are transformed to functionings valued.

restricted. This is evident with the HDI, which, although aimed at measuring human development, defined by the UNDP (1997:15) as "the process of enlarging people's choices[33]", was restricted to using the indicators life expectancy at birth, expected years of schooling, mean years of schooling and GNI per capita to measure the dimension indices.

The challenge of disaggregating data to estimate HC at a household or subgroup level is furthermore evident with the HCI estimate, which, when expanded to measure SES-HCI disaggregated by socioeconomic status, had to apply a reduced number of observed indicators and a reduced number of countries were studied.

This paper aims to aid future researchers when developing a model that reconciles the meaning and measurement of HC to estimated HC at a disaggregated level using a latent-variable approach and measure the inequality in its distribution. Such a HC estimate will make it possible to determine the relationship between HC and socioeconomic inequalities, which forms the basis of efficient provision of public goods towards expanding HC. The paper thus reviews the methodological development and empirical evidence of studies, which have worked towards obtaining the most accurate measure of HC at a disaggregated level, using the latent-variable approach, building upon the previous studies by Le *et al.* (2005); Oxley *et al.* (2008); Folloni and Vittadini (2010); and Slottje (2010). See Appendix 2.A, which provides, in the format of a table, summaries of the papers reviewed by considering the aims, method, data, variables, methodological strengths and weaknesses and findings of each. Furthermore, developing the latent-variable approach to estimate HC conceptualised by incorporating HCT and the capabilities approach is explored to ensure development beyond economic growth. Without accurate measurement and understanding of HC, precious resources may fail to expand human abilities towards citizens living a long and fulfilling life.

[33] Choices include to lead a long and happy life, education, a good standard of living, secure human rights and self-respect (UNDP, 1997).

2.2 Measuring HC as a latent variable

Dagum and Slottje's (2000) seminal study developed a method of measuring HC, which expanded on the initial attempts by Dagum (1994), Dagum and Vittadini (1996) and Di Bartolo (1999) to measure HC as a latent variable. Their key contribution in this study was the integration of the prospective method (developed by Farr (1853)), and the latent variable estimate method (using Wold's Partial Least Square (PLS) mode B method). This allowed them to develop what has become known as the latent-variable actuarial approach, which allows for calculating a monetary value of HH HC. Tao and Stinson (1997) had previously developed an integrated approach resolving problems inherent in the cost- and income-based methods, but it had not included the latent variable estimation method. The limitations of the new approach presented in Dagum and Slottje (2010) included that: HC was estimated solely as an unobserved composite variable of a set of observed formative indicators; the definition relied on headship; and suggested variables were excluded due to data limitations. Vittadini *et al.* (2003) noted that HC did not account for reflective or composite indicators to ensure it was fully consistent with the economic definition as per HCT. Vittadini *et al.* (2003) therefore recommended labour income as the reflective income, which restricts the measure of HC to economic outcomes. A further critique of the applied definition of HH HC was that it relies on headship to classify the HH, which is critiqued in the academic literature (Rogan, 2016), warning against using headship or a single reference person to classify HHs. See Appendix 2.A for the data sources and variables used in each of the studies reviewed.

Appendix 2.A highlights that the variables included in all the studies represent an extensive range of indicators of HC, and thus the latent-variable approach presents an opportunity for the capabilities approach to be included in the estimation of HC. For example, Dagum and Slottje (2000) included demographic factors as formative indicators and thus, HC is conceptualised in consideration of the presence of labour market segregation and discrimination. Therefore, following the capabilities approach, the model specified in Dagum and Slottje (2000) includes factors which affect the ability to convert the means to achieve, to the freedom to achieve functionings valued by specifying internal (for example gender and age) and external factors (for example parental socioeconomic status) (Chiappero-Martinetti and Sabadash, 2014). However, data limitations were evident and suggested variables including socioeconomic status of the parents, intelligence, ability, and genetic endowment of both the household head (HHH) and household spouse (HHS) were excluded given insufficient data. Dagum and Slottje (2000) estimated the average HH HC of households in 1983 to range from $238, 703 (cross-section average) to $364, 869 (life cycle average) (Dagum and Slottje 2000).

In a follow-up study, Dagum *et al.* (2003) applied the latent-variable actuarial approach and the same data source as Dagum and Slottje (2000) to measure HH HC. The results were then used in a recursive model to analyse the HC, income, and wealth determination distributions and derive the model's short- and long-term multiplier matrices. Data limitations were once again experienced. The estimated inequality measures, ranked from most unequal to least unequal were debt, wealth, HC and income, with calculated gini coefficients of 0.736, 0.636, 0.528 and 0.444, respectively. It was furthermore estimated that, in the short term, a one year increase in the schooling of HHH increased HH HC by $24861 versus $1124 for an additional year of full-time work. Thus, an extra year of schooling by HHH increased HC by 22 times more than a year of work experience. A further finding was that the additional contribution to HH HC of one year of schooling by the HHS was 4.6% higher than an additional year of schooling for the HHH. It is predicted that by expanding the estimation of HC to account for the direct and indirect values of human abilities, the contribution of indicators such as education would be multiplied, given it is potentially the means and the ends of development. An example is Wigley and Akkoyunlu-Wigley (2005) study, which identified to avoid underestimating the value of education, it should be estimated in terms of the capabilities to achieve functionings valued.

Vittadini *et al.* (2003) further developed the Dagum and Slottje (2000) method of measuring HC as a latent variable. Vittadini *et al.* (2003) highlighted numerous ways in which a latent variable like HC can be defined, which has implications for the statistical method by which it is measured. They noted that HC might be estimated using Factor Analysis if the latent HC is identified as a factor explaining the observed variance in the observed reflective indicators. However, when HC is defined as a combination of observed formative indicators, then the PLS method of analysis would be applicable. Vittadini *et al.* (2003) noted a third, more comprehensive option is to define the HC latent variable as a multidimensional construct, which is a linear combination of its formative indicators which best fit the reflective indicators in a path analysis model. In accordance with HCT, Vittadini *et al.* (2003) suggested household income as the only reflective indicator. Sen (1997) noted that although analysing HC focused on economic outcomes, an important part of development, it should be supplemented by recognising human beings as more than the means of production and the end of development.

An example is the inclusion of education as an observed formative indicator in the latent-variable approach (Vittadini *et al.*, 2003). Thus, although capturing education as part of HC's formation, education is a potential reflective representation of human abilities. To expand, Chiappero-Martinetti and Sabadash (2014) aimed to integrate HCT and the capabilities approach in the discussion on the value of education identifying that education acts as an end through acquired knowledge (enhanced self-esteem and an enriched understanding of the world) as a means to achieve wellbeing in the labour market (HCT) and as a conversion factor to convert knowledge to a long and healthy life. Indicators to capture the value of education should thus be extended to include for example type of employment (education as a means), education attainment and abilities (education as an end) and doing regular exercise, involvement in politics, human rights and religious groups (education as a conversion factor)[34]. Chiappero-Martinetti and Sabadash (2014) thus highlighted a broader vision of wellbeing should be accounted for, including intrinsic individual satisfaction (eg improved self-esteem) and externalities such as social progress.

[34] Indicators affecting conversion factors included parental education, age, gender and education attainment. See Table 9.1 in Chiappero-Martinetti and Sabadash (2014) for a comparison of variables used for measuring HC and capabilities.

Vittadini *et al.* (2003) used the same data source as Dagum and Slottje (2000) to calculate HH HC as a latent variable. Thereafter it was transformed into monetary values (using the avg HH HC calculation of $238 703 calculated in Dagum and Slottje [2000]). The distribution of HH HC from Vittadini *et al.*'s (2003) analysis revealed a high concentration of HH HC in the lower levels of HH income (using the monetary value). See Appendix 2.A for the numerous advantages of the method presented in Vittadini *et al.* (2003).

Foldvari and van Leeuwen (2005) relied on the earlier work by Tao and Stinson (1997) and Dagum and Slottje (2000), which proposed an integrated cost- and income-based estimation method to estimate the monetary value for national and per capita HC. Data from HH surveys in 1993, 1995 and 1997 in six Central and Eastern European countries was used, and the results were compared to Dagum and Slottje's (2000) HC estimates for the United States in 1982. Unlike Dagum and Slottje (2000), Foldvari and van Leeuwen (2005) could not estimate HC as a latent variable because their data sources did not contain adequate socioeconomic indicator data for latent variable estimation. Foldvari and van Leeuwen (2005) noted drawbacks of a latent-variable approach to measuring HC are the required sophisticated methods and high-quality data. Please see Appendix 2.A for a summary of these results. Instead, HC was measured as the present value of future earnings. Individual HC was not estimated but only average and aggregate HC. The results revealed that all six of the European countries had much lower HC accumulation than the U.S. Foldvari and van Leeuwen (2005) noted the importance of the research was partly that it was the start of a worldwide monetary comparable database of HC.

Further progress was made in measuring HC in 2007 when both Dagum *et al.* (2007) and Vittadini and Lovaglio (2007) expanded the research on the new method presented in Vittadini *et al.* (2003), which had proposed estimating HC as a standardised "unobserved multidimensional linear combination of formative indicators" which best fit the reflective indicators in a path analysis model. Both these new studies reiterated that using the PLS method to estimate HC (HC_{PLS}) (as had been done in Dagum and Slottje (2000)) would not be consistent with the economic definition of HC compatible with HCT; HC measured in this way would only be measured in terms of formative indicators. Reflective indicators measuring the return on investment would not be accounted for (Dagum *et al.,* 2007; Vittadini and Lovaglio, 2007). To address this defect, Dagum *et al.* (2007) explored both the formative and reflective indicators to be included in the estimation of HH HC, consistent with a meaningful definition of HC as per HCT. Again, the HH HC was classified by the HHH and HHS characteristics. Dagum *et al.* (2007) estimated HH HC using 2004 US data. Vittadini and Lovaglio (2007), however, used the same method and data as Dagum and Slottje (2000) to compare the impact of the new approach to estimating HH HC, as well as changes in the distribution of HC. Vittadini and Lovaglio's (2007) findings are discussed first below; thereafter the results from Dagum *et al.* (2007) are examined.

Vittadini and Lovaglio (2007) found that both the HHH and HHS education have highly positive and significant impacts on HH HC in both. Some of the main differences between their results and those of Dagum and Slottje (2000) illustrated that the newer method was more consistent with the economic definition of HC proposed by HCT. Firstly, HHH age, which had shown an unexpected negative relationship in the Dagum and Slottje (2000) study, revealed the expected significant and positive relationship. This new result is in line with the expectation that as age and work experience increase, both HC and income should increase. Secondly, geographic region was shown to have a small but positive weight with the new method, compared with its negative and significant weighting in the Dagum and Slottje (2000) study. Again, this new finding was in line with the expectation that living in more affluent regions should be positively correlated with HC. Interpreting the results in line with the capabilities approach, the geographical area (providing a set of labour market opportunities) could represent a conversion factor (when considering education as a means to development), impacting the conversion from endowments, such as education inputs, to the freedom to achieve functions valued (Chiappero-Martinetti and Sabdash, 2014).

The results from Dagum and Slottje (2000) and Vittadini and Lovaglio (2007) were compared by estimating their income-generating functions using linear regressions of the different HC estimates on HH earned income (Vittadini and Lovaglio 2007). The results highlighted the superiority of the new method, which had higher R^2 and F-statistics and a lower Root mean square error (RMSE). Next, the actuarial mathematical approach from the Dagum and Slottje (2000) study was applied to the new latent variable estimates to compare the monetary values and distribution of the two methods (HC_{PLS} versus $\widehat{HC}$). The empirical distributions were significantly different, which has important consequences for future attempts to analyse HH HC inequality. Both methods reveal highly unequal distributions of HH HC, HC_{PLS} versus $\widehat{HC}$, with the new method revealing an even more unequal distribution than the former. Please see Vittadini and Lovaglio (2007) for a graphical illustration of the distribution results.

Dagum *et al.* (2007) applied the newly proposed method to more recent data, estimating short- and long-term multipliers through a recursive model to study causation, in line with a recommendation made in the Vittadini and Lovaglio (2007) study. To remain consistent with the economic definition of HH HC, as identified by Dagum and Slottje (2000), Dagum *et al.* (2007) methodically selected the indicators for the estimation of microeconomic HH HC. Accordingly, they analysed all the indicators used in the retrospective, prospective and education-based approaches to choose the most accurate indicators to inform HH HC. The new method was subsequently applied using the 2004 Survey of Customer Finances (SCF) dataset to measure the HH HC in the US. However, they note that no adult skills or parental characteristics were available for the HHH and HHS in the SCF database. Dagum *et al.* (2007) suggested these be included amongst the formative indicators, and their absence was thus a limitation of the new estimation of HC. In addition, the only reflective indicator that could be included, given the available data, was earned income (net wealth).

The weight estimates of the formative indicators suggested that the indicators related to education and job training have the most significant impact on HC formation, consistent with expectations. These indicators included years of schooling, years of full-time work, job status of HHH and type of occupation. Regarding demographic indicators, the gender of the HHH had the largest weight (illustrating the existence of gender inequality in the US) and, within the family, marital status had a higher weighting than the number of children.

Once the formative indicators had been weighted, the actuarial method introduced by Dagum and Slottje (2000) was applied to obtain a monetary value for HH HC. Thereafter the causal recursive model for HC was used. This allowed for the short and long-term multiplier matrices to be estimated to assess the causation of direct and total effects of the endogenous variables. Compared to the Dagum *et al.* (2003) study, a broader range of variables was included. Labour market formative indicators, which were found to have a substantial impact on monetary HH HC included type of occupation, sector and job status, which had a strongly positively related impact, and part-time employment, which had a negative impact on monetary HH HC (Dagum *et al.,* 2007). Amongst the educational formative indicators, years of schooling was noted to play a significant role in determining monetary HC. Of the exogenous formative indicators, gender had a more weighted impact on estimating standardised HH HC than the age and race of the HHH. However, when accounting for labour market factors, the monetary HH HC impacts of race were greater than gender, indicating racial discrimination in the labour market (Dagum *et al.,* 2007). As Vittadini and Lovaglio (2007) and Dagum *et al.* (2007) identified, aligning the definition of HC with the indicators is required to supplement the analysis of public policy towards enabling citizens the freedom to live lives they value.

Lovaglio (2008) further developed the method of measuring HC by studying HC as a bidimensional latent variable compared to a unidimensional latent variable presented in the previous studies. Concomitant indicators were added, following the path analysis rules, in a measurement and structural model. Please refer to Lovaglio (2008) for the development of the equations used to estimate HC. Consistent with HCT, Dagum *et al.* (2007) and Vittadini and Lovaglio (2007) had both noted that education and work experience are expected to be the most significant indicators of HC. In line with this expectation, Lovaglio (2008) defined HC as a latent variable composed of the principal determinants, namely education (EduHC) and work experience (JobHC) of the HHH and HHS, which contributed to causing HH earned income and capital income (reflective indicators). Lovaglio (2008) noted the need to control for the indirect impact of concomitant indicators (for example wealth), which are observed exogenous variables directly linked with the reflective indicators of the latent variable (LV) HC (eg earned income) but are not embedded in the formative block (eg education). The importance of education and work experience as sources of lifetime income are thus identified. The inclusion of concomitant indicators in the estimation of HC was one of the main differences between this study and the studies by Dagum *et al.* (2007) and Vittadini and Lovaglio (2007).

Lovaglio (2008) firstly calculated HC factors scores from the measurement model and converted the scores to monetary values, using an extended method (to account for the bidimensional nature of HC) of the initial actuarial approach developed in Dagum and Slottje (2000). Thereafter, Lovaglio (2008) applied a structural model consisting of three simultaneous linear equations to specify the causal links between earned income, capital income, JobHC, EduHC, wealth and debt, by generalising Dagum's recursive model (Dagum, 1994). Despite the developments of this study, it is evident that the analysis of HC is restricted to economic outcomes; however, the ability of the formative-reflective model to be expanded to a bidimensional construct highlights the potential of the capabilities approach to be included in the conceptualisation and measurement of HC.

In this new study, Lovaglio (2008) focused on Italian HH HC using data from the 2000 Italian Survey on Income and Wealth. Dagum *et al.* (2007) had identified the importance of including social and demographic variables as necessary indicators, but these factors were allocated between JobHC and EduHC in Lovaglio (2008) without explanation or empirical support. Dagum *et al.* (2007) suggested the demographic indicators included geographical region, age, ethnicity, gender, and other domestic characteristics, including the number of children. Lovaglio (2008) allocated age and geographical area to JobHC and the number of children and gender to EduHC. The allocation of gender to EduHC is controversial, given that gender does not consistently impact educational attainment (for example, in several countries, including South Africa with high levels of inequality, the learning adjusted years of the education component of the HCI is higher for females than males (World Bank, 2018a)). However, gender does impact labour market outcomes, particularly given the motherhood penalty, which has been highlighted as significant by researchers in different countries (Casale and Posel, 2005; Posel and van der Stoep, 2008; and Magadla *et al.,* 2019).

Ethnicity was furthermore not included in the new model presented in Lovaglio (2008). This reiterates a potential shortfall of the latent-variable approach to measuring HC (which can be extended to any approach measuring HC), namely, that a lack of a theoretically explicit definition of HC results in the absence of clearly prescribed observed variables (Costa, 2006). Lovaglio (2008) highlighted the importance of developing a local model which accounts for context specific indicators. Thus, the choice to include and allocate social and demographic factors should be based on empirical evidence and within the specified theoretical framework. This is relevant when selecting the indicators to measure HC when supplementing HCT with the capabilities approach to account for both the direct and indirect values of human abilities.

An example of the importance of context and the context within which the data is collected is evident when critiquing the use of characteristics of the HHH and HHS to classify the HH (Rogan, 2016). Survey respondents may interpret the meaning of headship differently depending on culture, age, race, household structure, etc. (Rogan, 2016). For example, in many developing countries, extended families may reside in the same household, which complicates determining a single HHH. Furthermore, household surveys may assign headship differently or arbitrarily (Chant, 2007).

The results from Lovaglio (2008) showed that for Italian HH JobHC in 2000, the number of years of full-time work of HHH, age of entry into the labour market of HHH, and HHH age, were all statistically significant, with coefficients of 294.308, -493.030 and 161.364 respectively. Regarding EduHC, years of schooling of HHH and HHS indicated that the education of the spouse (co-efficient 707.980) contributed more to EduHC than that of the HHH (co-efficient 687.528). The contribution of the type of university degree was negligible, indicating that possession of a university degree, measured through additional years of education, which had a coefficient of 606.57 for HHH and 707.98 for HHS, is more significant than the type of degree. The impact of intergenerational EduHC was evident in that the schooling of the HHH father, the parents of the spouse and particularly the level of education of the HHH's mother all contributed to informing EduHC. Gender inequality was furthermore evident in the formation of EduHC in that the coefficient was 1456.621 for the dummy gender variable (Lovaglio, 2008).

The earned income, EduHC and JobHC (presented from smallest to largest) mean values were calculated as €17 525, €82 659 and €128 881, respectively. Inequality for JobHC was lowest, with a gini ratio of 0.327, whereas EduHC recorded a gini ratio of 0.593. The inequality for earned income was similar to that of JobHC with a gini ratio of 0.353. HH debt had by far the largest gini ratio of 0.924.

Following on from estimating the monetary value of HC using the measurement model, the causal model specified by Lovaglio (2008), which generalised Dagum's recursive model, was used to estimate the relationships between EduHC, JobHC, wealth, income and debt. Both EduHC and JobHC were found to strongly and positively contribute to the return of HC to earned income. When tested for significance, JobHC proved to have a more significant impact than EduHC, with JobHC, however, strongly dependent on EduHC (Lovaglio, 2008). Wealth (and to a lesser extent JobHC) were found to be essential sources of capital income. EduHC and capital income were, surprisingly, negatively related. From the results, Lovaglio (2008) provided support for Becker's (1964) suggestion that more risk-averse individuals would focus on investing in education (EduHC), given its positive relationship with earned income, as opposed to investing in physical stock.

Lovaglio (2008) compared the calculated bidimensional estimate of HH HC for Italy in 2000 with the unidimensional calculation for the US in 2004 (Dagum *et al.,* 2007) and found that Italy has a HC measure (€187 440) approximately a quarter of that of the US ($852 533). However, HH HC inequality was lower in Italy than the US, with gini ratios of 0.52 and 0.656, respectively. Lovaglio (2008) noted that similar differences were evident when comparing the distribution of HH earned income with the US mean earned income i.e. HH earned income in the US was 2.5 times the size of that of Italy, and earned income inequality in the US was 40% greater than in Italy. It was reasoned that since market demand for the type and amount of personal HC determined personal income, increasing EduHC and reducing the inequality of its distribution would reduce inequality in Italy's personal income levels.

In a later study, Lovaglio (2010) estimated HC at the worker level using Italian regional data in a Reduced Rank Regression (RRR) framework. The Lombardy regional administration archive, known as 'Employment Centres of the Province of Milan', which collected data concerning workers in the private sector of the Milan, was used to source the required data. The data was described as being of good quality, thus allowing for an enriched and broadened analysis. Furthermore, the detailed data allowed for Lovaglio (2010) to conduct a longitudinal study estimating changes in HC at the worker level over the period 2000 – 2004. The longitudinal analysis aimed to explore how workers' personal characteristics, particularly education, impacted on growth in their earned income.

In the study, HC was measured as a multidimensional (as opposed to a uni- or bidimensional in previous studies) latent variable (Ξ). This was illustrated in a path-diagram model for the individual worker, with the reflective indicator being restricted to earned income (Y). Please see Lovaglio (2010) for an illustration of the path diagram. Concomitant indicators (W_1) were accounted for in the causal model. These included sociodemographic indicators, which impacted the reflective indicator but were not part of HC investment formative indicators. They also impacted additional concomitant indicators, representing opportunity factors of HC but not HC investment formative indicators. The HC investment formative indicators (Z) that were used were education and days of work and training. Concomitant formative indicators were gender, age, marital status, number of children, type of contract, industry, occupation, gender and nationality (W_1). Parents' socioeconomic status and wealth data were not available and thus were not included in the analysis. Additional data not available included measures of education quality and the characteristics of the educational institution attended or of training on the job.

The Lovaglio (2010) study allowed for demographic indicators to be allocated more accurately as concomitant indicators instead of as components of the bidimensional HC measures as suggested in Lovaglio (2008). It also provides the opportunity to measure HH HC by firstly measuring individual worker level HC and thereafter aggregating the estimates of workers in the HH to calculate a HH HC value. Therefore, the definition and data of the HH HC are not restricted to that of the HHH and HHS. HC is thus initially estimated at an individual level, within the context of the available data, and then HH HC per capita calculated by aggregating the individual member's HC estimates.

Lovaglio (2010) showed that estimated HC and concomitant indicators explained 56% of the variation in earned income. Two-thirds were explained by the concomitant indicators, namely, and in order of importance, gender, age and occupation type. Years of schooling had a 57% weighting towards HC formation, with days of fulltime work and days of training having weightings of 31.8% and 11.1%, respectively. The monetary value of HC was calculated using the actuarial method of Dagum and Slottje (2000); it was found, as expected, that years of schooling had the largest impact on monetary HC. Furthermore, it was noted that the gini coefficient for HC (0.501) was greater than that of income (0.389).

The longitudinal study results showed that income variability mainly was accounted for by between worker differences (78%), whereas time only accounted for 8% of the variation (Lovaglio, 2010). Individuals with a university degree, and those with contractual evolution towards a permanent contract, experienced the fastest income growth. Gender discrimination had the most significant impact on earnings, with males aged 41 – 50 years, with a university degree and a permanent contract, experiencing the highest remuneration. Gender differences in income growth were particularly evident in the most educated group (university graduates) and the most productive age group (20 – 40 years).

The methods of calculating individual HC were extended by Lovaglio and Folloni (2011), Lovaglio and Vittadini (2014) and Lovaglio *et al.* (2016), using data on graduates in Milan. Lovaglio and Folloni (2011) proposed a class of models for estimating individual HC, which allowed for various relationships among latent variables and endogenous indicators to be specified, particularly to allow for covariate effects. Lovaglio and Vittadini (2014) proposed a new method of Generalised Redundancy Analysis (GRA) to fit Structural Equation Models (SEM) to measure latent variables such as HC, allowing for a variety of relationships among composites, endogenous variables, and external covariates to be specified and fitted. HC was measured as a bidimensional variable consisting of EduHC and JobHC as had been introduced in Lovaglio (2008). A limitation of the study was that the approach was descriptive without inferential aspects, which was noted as of growing interest when applied to the population as a whole (Lovaglio and Vittadini, 2014).

With regards to concomitant indicators, it was concluded that the impact of family background on HC is indirect and is accounted for in education choices and achievements. Lovaglio *et al.* (2016) estimated graduates' HC within a realistic structural model accounting for possible effects of external covariates. The GRA approach was adopted, as had been suggested in Lovaglio and Vittadini (2014). Data limitation and selection bias were among the limitations raised. The analysis concluded that there was evidence that graduation employment positively affected labour market outcomes (i.e. earnings) after graduation. However, it was found that graduates with higher academic ability (in times of economic downturn and poorer labour market prospects) invested further in education (i.e. postgraduate studies), postponing entrance into the labour market.

Although these studies provide methodological developments, they included detailed individual level variables in the path model, which are not available through HH surveys or census datasets. Thus, not all the variables included in the path model are relevant to calculating HH HC when aggregating the HC variables of all economically active household individuals. The methodological developments reveal how multiple variables and their relationships could be included when estimating

HC and thus potentially allow for an estimate of HC reconciled with the meaning of a multidimensional HC to account for both the indirect and direct values of human abilities. However, the practical application of the methodological developments and data availability are expected to remain limiting factors. Nevertheless, Chiappero-Martinetti and Sabadash (2014) identified the HCT and capabilities approach paradigms have the potential to complement each other when valuing education by extending the range of variables.

Finally, Fattore *et al.* (2018) took a different theoretical approach to measure a latent variable, compared to the prior structural equation methods reviewed, by providing a new least-squares approach. Numerous advantages were noted for this approach. Firstly, it effectively extracted and demonstrated improved performance in formative-reflective models measuring latent variables, compared to PLS path modelling (PLS-PM) and SEM. The paper applied the method to a real-world example constructing the latent variables of agricultural inequality, industrial development and political instability in a path model where the first two were expected to cause the latter. The results from the SEM, PLS-PM and the suggested new least-squares approach were compared. The first two showed more discrepancies (Fattore *et al.,* 2018). In addition, the new least-squares approach accounted for not only descriptive but also predictive and causal analysis (Fattore *et al.*, 2018). The authors also made software for R package available ('pathmod'), which allows for easy application of the method.

Thus, this method is a viable option for measuring HH HC where the model is consistent with the methodology. Furthermore, an accurate definition of HC and context and data-appropriate variables, as suggested by authors including Sen (1997), Dagum *et al.* (2007), Lovaglio and Vittadini (2007), Lovaglio (2008) and Chiappero-Martinetti and Sabadash (2014), are required to estimate the direct and indirect values of human abilities. Different indicators will be more relevant in studies in different countries or regions. For example, Lovaglio (2008) identified in Italy, low social mobility could result in education disadvantage perpetuated over generations, and it would thus be relevant to include parental educational characteristics. In addition, data availability determines what can or cannot be included in any model measuring HC as a latent variable. Conceptualising and measuring HC by considering HCT and the capabilities approach would require micro-level data and thus a challenge of incorporating both theoretical frameworks.

2.3 Conclusion

Sustainable development is not possible through relying solely on growth (Evans 2010) and non-human capital. Sen's capabilities approach specifies that human capabilities are both the goal and the means of development. Sen (1999) furthermore acknowledged that capabilities cannot be reduced to one good. Thus, measuring HC requires investigation beyond education. Researchers should explore statistically how to define and measure HC as a latent construct. Reconciling the meaning and measurement of HH HC estimates enables the causal relationships between HC, income, wealth and the related inequality of each, to be determined. Thus, it will be possible to ensure efficient investments in HC, required at the micro- and macro-level, will reduce identified inequalities.

However, when considering using a latent-variable approach to measuring HH HC and studying HC inequality and the consequences, locating an appropriate method may be overwhelming, given the statistical complexity of measuring non-observed constructs. This paper thus discusses the development of the methods and presents a table summarising each as a contribution towards assisting researchers in constructing a model to estimate HH HC when defining HC as a latent variable. Considerations of incorporating the capabilities approach to measuring HC using the latent-variable approach are furthermore highlighted.

In developing a model measuring HC at the HH level, Vittadini and Lovaglio (2007), Dagum *et al.* (2007), Lovaglio (2008) and Lovaglio (2010) are useful additions for identifying the variables to be included. It is recognised that the definition and indicators to be included in the measure of HC as a latent variable are not definitive and thus indicators included in a measure of HH HC should be refined as per data availability, specific context and theoretical framework. In accordance with the HCT and the studies reviewed the variables considered should broadly fall under the headings of education, work experience, indicators linked to career progression and discrimination and finally income as a reflective indicator of HC. Sociodemographic indicators should also consider using variables concerning the parents of individuals, given the intergenerational transfer of capital and welfare (Lovaglio, 2010). The reflective indicator suggested in the studies reviewed was consistently income, which is a potential entry point for the capabilities approach. Thus, to supplement HC analysis as a determinant of socioeconomic inequality, HC should be conceptualised as the direct and indirect values of human abilities and thus require a reconsideration of the indicators and the path model specified according to empirical evidence HCT and the capabilities approach. As suggested by Chiappero-Martinetti and Sabadash (2014), variables selected should consider observed indicators representing the wellbeing process in which human endowments (private and public) are transformed to a set of functionings valued (achievements). Within the process the capabilities set, which mediates the transformation, representing the real options available should be accounted for. Furthermore, internal and external conditions, which affect the transformation from endowments to capabilities, as well as the choices, transforming the capabilities to functionings valued, should be recognised. Although, factors within the wellbeing process may be resistant to measurement and require micro-level data, supplementing estimating HC beyond income outcomes and broadening the vision of wellbeing will supplement the analysis of HC as a determinant of socioeconomic inequalities. Reviewing the developments of the latent-variable approach has highlighted the potential for estimating HC as a multidimensional construct and the considerations required to measure HC by supplementing HCT with the capabilities approach.

An example of an indicator that should be considered and which has not been discussed in much detail in the LV studies reviewed, is health (Grossman, 2000 and Kraay, 2018). Grossman (2000) argued that unlike some of the outputs of HH production function, which either directly enter the utility function or are related to determining earnings or wealth in a life cycle framework, health uniquely does both. As a dimension of HC, health is inherited as an initial endowment, but deteriorates after time, at an increasing rate in the later stages of the life cycle. But it can also be increased through investment. Given Lovaglio's (2008) finding that EduHC decreased as a proportion of HC overtime, and JobHC increased, it is presumed that health HC would also decrease with age until at death when it would be zero. Applying a multidimensional measure of HC should thus aim to include health as a principal indicator of HC, similar to education and work experience. The index measures of HC discussed by Liu and Fraumeni (2021) include health indicators, for example the HCI includes a health component with adult survival and under-5 stunting rates included as proxies for health (Kraay, 2018). Although these measures are national-level measures, different household surveys, which include a section focused on health, should allow for potential indicators of health at the microeconomic level.

A challenge when investing in HC is that the impact thereof may only result in returns in the distant future. Social returns may also be higher than private returns, restricting HC estimates to income as the reflection of HC leading to under-investment (Evans, 2010). It is hoped that this paper has highlighted issues which need to be considered by researchers when estimating and analysing HC in order to ensure efficient and equitable investment in HC.

2.4 Appendices

Appendix 2.A: A summary of studies reviewed.

Title/Authors	Aim	Method	Variables & Data source (Directly from each paper)	Definitions/ Assumptions/limitations	Key advantage/ contribution	Main findings
A new method to estimate the level & distribution of household human capital with application. Dagum & Slottje (2000)	1) Identify indicators to estimate latent variable (LV) HC 2) Method to transfer from dimensionless HC estimation to monetary value 3) Method estimating avg HC to estimate HC distribution 4) Apply estimation methods to measure US 1983 HH HC & distribution	Estimate HC as a standardized LV using Wold's PLS mode B method. Convert HC into a monetary value using actuarial mathematical method.	Data: 1983 US FRB sample survey of consumer finances Variables: x1 =HHH, age x2 = region x3 =HHH, marital status x4 = HHS, gender x5 =HHH, years of schooling x6 = HHS, years of schooling x7 = number of children x8 =HHH, years of full-time work x9 = HHS, years of full-time work x10 =HH total wealth x11 = HH total debts	HC estimated as unobserved composite variable of set of observed indicators F (formative indicators); Not fully consistent with economic HC definition; Variables suggested but data not available for: socioeconomic status of the parents of the HHH, socioeconomic status of parents of HHS, intelligence, ability & genetic endowment of the HHH & HHS; Assume given pattern of productivity increase by age when converting to monetary values; Estimates HC as zerodimensional.	New method of measuring HC as a LV using Wolds PLS; Converts LV HC to monetary value using developed latent actuarial approach	Model estimating HC results statistically significant at 1% level. R squared of 0.618; Independent variable highest coefficient years of schooling of HHS then HHH; Variable lowest coefficient HHH years of experience. Gini ratio 0.528, smaller than Gini ratio for total wealth (*G*_0.636) & for net wealth (*G*_0.681); BUT greater than the income inequality (*G*_0.444), estimated for same survey data (Dagum, 1994). Estimated avg HH HC ranged $238,703 (cross-section avg) to $364,869 (life cycle avg).
A Microeconomic recursive model of HC, Income & wealth determination: Specification & Estimation. Dagum, Vittadini, Lovaglio & Costa 2003	Explain HH HC, income & wealth; 1) determination (establish a pattern of causation for 17 endogenous variables) 2) distributions; & 3) derive the model short & long term multiplier matrices.	Recursive model for determining causation between 17 variables; HC variable 16 & estimated using Dagum & Slottje (2000) method; Recursive model for quantitative endogenous variables estimated using OLS & qualitative endogenous using logit approach to parameter estimation; Dagum model fitted to determine distribution of	Data: 1983 US FRB sample survey of consumer finances a) Exogenous variables x1 = age HHH; x2 = gender HHH; x3 = race HHH; x4 = region residence; x5 = marital status HHH; x6 = age HHS; x7 = gender HHS; b)Endogenous variables: z1 = yrs schooling HHH; z2 = yrs schooling HHS; z3 = number of children; z4 = yrs full-time work HHH; z5 = yrs not full-time work HHH; z6 = years full-time work HHS; z7 = years not full-time work HHS; z8 = job status HHH; z9 = occupation HHH; z10 = industry HHH; z11 = job status HHS; z12 =	Assumptions same as Dagum & Slottje (2000); Predetermined variables not included – lack of data: lagged values income & wealth HH, & particularly, variables of parents of HHH & HHS: income & wealth at time of birth & time of marriage of HHH & HHS, years of schooling & place of residence at time HHH & HHS	For 16 of 17 variables, model shows extreme goodness of fit' Policy implications from results regarding HC; income, wealth, growth, development. Short & long term outcomes.	Inequality calculated from largest to smallest, debt, wealth, HC & income; HC results from recursive model: Marginal increases of HC, resulting from 1 year of schooling increase of HHH & HHS$24861 & $25999, respectively. One year increase of full-time work of HHH increases HC by $1124; in ST, one dollar increase of total wealth (k) increases HC by $0.029, one dollar increase of total debt (D) increases HC by $0.621, consistent with economic units policy of indebtedness to further accumulate HC. Debt requires a shorter time lag than K to have

		income, wealth, debt & HC.	occupation HHS; z13 = industry of HHS; z14 = k = household total wealth; z15 = d = household total debt;z16 = HH HC; z17 = y = HH income.	completed high school.		impact on HC; (vii) one dollar increase of K & HC increases Income by $0.571 & $0.30 respectively. HC decreases by age of HHH - however not consistent with all other research.
The measurement for the estimation of the distribution of HC from sample survey on income & wealth. Vittadini, Dagum, Lovaglio & Costa (2003)	Proposes to measure distribution of HC. New method to measure HC as a LV in a path analysis model as opposed to factor model & PLS model.	Further develops Dagum & Slottje (2000) by measuring HC (as multidimensional) as a LV in a path analysis model (subset of Structural Equation Modelling (SEM)) by considering both formative & reflective indicators of HC. Uses path analysis as better than factor model & PLS model. Uses Dagum & Slottje (2000) actuarial method to convert HC to monetary values & distribution.	Data: 1983 US FRB sample survey of consumer finances; Indirect indicators: x2=HHH Gender; x3 = HHH Race; x6 = HHS Age; y5 = HHH Years Not Full-Time Work; y7 = HHS Years Not Full-Time Work; y8 = HHH Job Status; y9 = HHH Occupation; y10= HHH Industry; y11= HHS Job Status; y12= HHS Occupation; y13 = HHS Industry; Formative indicators: x1 = HHH Age; x4= Region; x5 = HHH Marital Status; x7= HHS Gender; y1= HHH Years of Schooling; y2 = HHS Years of Schooling; y3 = Number of Children; y4 = HHH Years of Full-Time Work; y6= HHS Years of Full-Time Work; y14= Household Total Wealth; y15 = Household Total Debts Reflective indicator: Income.	As per Dagum *et al.* (2003) & Dagum & Slottje (2000) lack of data: lagged values income & wealth HH, & particularly, variables of the parents of HHH & HHS: income & wealth at the time of birth & at the time of marriage of HHH & HHS, years of schooling & place of residence at the time HHH & HHS completed their high school. Does not account for concomitant indicators.	"1) Estimates HC from manifest variables (MV), consistent with causal relations, avoiding treating formative as reflective & vice versa. 2) Parameters are estimated in a causal model framework "3) Approach is nonparametric. 4) In case of many dependent reflective indicators can be generalized by…Redundancy Analysis (Tso, 1981), proposed in PLS Path modelilng (Tenenhaus, 1995; Lovaglio, 2001)." Vittadini *et al.* 2003	Establ zzished several advantages on the model as seen under advantages. Use new method to measure distribution of HC as a LV & in monetary value using the actuarial approach developed in Dagum & Slottje (2000).
An Estimation of the Human Capital Stock in Eastern & Central Europe. Foldvari & van Leeuwen (2005)	Measure average & aggregate HC using Dagum & Slottje (2000) (excluding LV part) in 6 Central & Eastern European countries compared to US HH HC (1983 data)	Dagum & Slottje (2000) method – however not LV estimation as not sufficient data. Heckman selection model to correct for bias in Dagum & Slottje (2000) model; HC estimated as present value of future earnings.	Available household surveys for 1993, 1995, & 1997 US HH estimation by Dagum & Slottje (2000) which used 1983 US FRB sample survey of consumer finances. Variables: After tax income, age, gender, level of education, HH size	Could not conduct SEM because of lack of socioeconomic variables. Only focused on average & aggregate HC. Not HH or individual.	Comparison on numerous countries	Slovakia (in 1993) - country faired best in terms of per capita HC in comparison to the US. Slovakia experienced 32.86% of the US 1982 per capita HC in 2000 US dollars, country laggin most behind the US was Estonia in 1995 with 20.77% of the US 1982 per capita HC (Foldvari & van Leeuwen, 2005). The European countries studied further lagged significantly behind New Zealand, which Foldvari & van Leeuwen (2005) found was 71.56% in 1996 of the US per capital 1982 HC estimate.
FORMATIVE INDICATORS & EFFECTS OF A	Introduce an improved method of estimating	Uses path analysis model introduced in	US 2004 Survey of consumer finances. Variables selected in accordance with	No adult skills or parental characteristic indicators	Same as Vittadini *et al.* (2003) 'new' method.	Gender more of an impact when looking at weights than race which has a greater impact

CAUSAL MODEL FOR HOUSEHOLD HUMAN CAPITAL WITH APPLICATION Dagum, Vittadini, and Lovaglio (2007)	HH HC as LV in agreement with economic definition of HC. Estimate monetary HC to obtain multipliers measuring effects of variables	Vittadini *et al.* (2003). Recursive model for multipliers Dagum *et al.* (2003) Actuarial method for monetary value Dagum & Slottje (2000).	economic definition of HC. Reflective indicator household income. Formative indicators HHH years of schooling; HHS years of schooling; HH debt; Edu loan; HHH gender; HHH age; HHS age; HHH race HHH marital status; number of children HHH job status; HHS job status; HHH type of occupant ion; HHS type of occupation; HHH years of full-time job; HHS years of full-time job; HHH part-time job; HHS part-time job; HHH sector; HHS sector; HHH age starting job; HHS age starting job	(variables) useful for the estimation Obtained; Does not account for concomitant indicators & only 1 reflective indicator.	Method allows for policy recommendations specifically with regards to link between HC & wealth & income inequality.	when looking at coefficients which account for labour market. Job status & education both having large impacts. Age having a negative impact on monetary HC. HC monetary mean for US households 2004 is $852,533, median = $982,401, Gini ratio is 0.656.
Evaluation of the Dagum-Slottje method to estimate household HC. Vittadini & Lovaglio (2007)	Introduce improved method of estimating HH HC – same as Dagum, Vittadini, and Lovaglio. (2007) & then compare results with Dagum & Slottje (2000) method using same data.	Uses path analysis model introduced in Vittadini *et al.* 2003; To compare use each method & same data; HC obtained by a linear combination of mixed formative indicators F that best fits reflective indicator (y = earned HH income).	Data: 1983 US FRB sample survey of consumer finances Same as Dagum & Slottje (2000) & Vittadini *et al.* (2003). Formative indicators same as Dagum & Slottje (2000). For Vittadini *et al.* (2003) formative indicators are extended according to the economic definition of HC. Variables: x1 =HHH, age; x2 = region; x3 =HHH, marital status; x4 = HHS, gender; x5 =HHH, years of schooling; x6 = HHS, years of schooling; x7 = number of children; x8 =HHH, years of full-time work; x9 = HHS, years of full-time work; x10 = household total wealth; x11 = household total debts	Analysis with more relevant variables & recent data would improve. Partly been done in Dagum, Vittadini, and Lovaglio.(2007) . Suggests recursive model to test for multipliers. Does not account for concomitant indicators & only 1 reflective indicator.	Same as Vittadini *et al* 2003. Improvement on measuring HC compared to PLS.	The new method producers results are more in line with economic definition of HC compared to Dagum-Slottje method; Schooling still a major impact, age & region have particularly different results from using the different methods. Also HC distribution results different.
Process of accummulation of Italian Human Capital. Lovaglio (2008)	Measures HC by composing HC as bidimensional (Education HC & Work experience HC) & reflected in earned & capital income. Generalize Dagum's recursive model & proposed a simultaneous equation model to estimate relationship	Use: Structural equations specified in Path Diagram to estimate EduHC & WorkHC as the best linear combination of formative indicators that best fit reflective indicators Y (earned & capital income) nett of direct effects of Wealth & net the effects of (opposite latent dimension) formative indicators to the contribution of Y.	Data source: 2000 Italian Survey on Income & Wealth. Variables: Z matrix observed variables of education (EducationHC); X matrix observed variables for Work Experience HC (JobHC); Y is the block of endogenous indicators (earnings); & W1 matrix contains Wealth indicators	Cannot be used as a single global model but sensitivity analysis needed.	Allows for concomitant indicators which previous methods do not. Measures two different dimensions of HC. Adding monetary value & recursive model allow for designing more effective policies.	Monetary estimated value of HC & HC inequality in Italy about a quarter of US, being €187440 & $852533 respectively & gini coefficient 0.52 & 0.656 respectively. Importance of EduHC in generating earned income is future EduHC reduces compared to JobHC.

	between earned income, capital income, JobHC, EduHC, wealth & debt.	Qualitative variables transformed into quantitative using ALSOS.				
The estimation of HC by administrative achives in a static & longitudinal perspective: The case of Milan. Lovaglio (2010)	Estimation of HC as a LV at the worker level in Milan, Italy. Identify correlation & causality between variables & estimate distribution.	"The methodology has estimated HC as the rank-one best linear combination (in an RRR framework of their formative indicators that best fits their reflective indicators (Y), net to the direct effects of concomitant indicators to Y, & net to the spurious effects of concomitant indicators on both HC & Y" (Lovaglio, 2010). Static & longitudinal analysis. Dagum & Slottje (2000) actuarial method to calculate monetary HC.	Data: Lombardy region administrative archive data on 'Employment Centers of the Province of Milan', & administrative flows data collecting mandatory workers' individual income tax returns, filed with National Internal Revenue Service. Reflective indicator: earned Y Formative indicators: years of schooling, days of full-time work in the entire period of observation (2000–2004) & days of training in the period Concomitant indicators: gender, age, marital status, number of children, type of contract, industry, type of occupation & nationality	Missing concomitant variables: parents' socio-economic status (wellbeing, workers health status, etc.) & wealth of origin household. Lack of data on worker career prior 2000. Use of administrative archives is related to the problem of self-selection & endogeneity of labour market participation. To replicate method to investigate another region need detailed microeconomic data.	Allows for impact of concomitant indicators to be accounted for. Worker level HC measured versus national & HH.	Italian job market marked significantly by inequality, gender, & when investigating processes of school to work transitions, finds weak incidence of educational impact on earned income overtime. Earned income impacted by HC but also by discrimination & career progression factors such as gender, age & type of occupation vs education/training dimensions. Longitudinal analysis 78% of total income variability is between workers & only 8% to variation overtime of individual workers.
Human Capital Measurement: A survey. Folloni & Vittadini (2010)	Reviews previous means of measuring HC & proposes new method of measuring HC as a function of formative indicators & reflective indicators (earned HH income). Previously used in Dagum, Vittadini, and Lovaglio (2007), Vittadini & Lovaglio (2007)	Literature study	New method suggests variables to be used. Formative indicators for HC eg educational attainment i.e. schooling & training while working; work experience; non-cognitive skills; parental & family characteristics. Reflective indicator eg earned HH income.	HC defined as a LV & thus measured as such. Need methods for missing data.	New improved method specified for measuring HC as a LV.	New method: HC estimated as linear combination of formative indicators that best fits earned HH income (Dagum *et al.*, 2007 & Vittadini & Lovaglio, 2007). Case of many dependent indicators, method generalized in PLS path modelling framework by means of reduced rank regression models & redundancy analysis (which has since been done in Lovaglio & Vittadini (2014), & Lovaglio *et al.* (2016). Further research suggested. Categorical indicators transformed by ALSOS in multiple regression model. Global model accurate? Need a local model. Using new method, Dagum's recursive model can explain household HC, wealth & income determinations & distributions to analyse HC policies.

						Lovaglio (2008/2010) method to add concomitant indicators.
Human Capital Measurement: Theory & Practise. Slottje (2010)	Literature review of HC measurement methods building on work of Le *et al.* (2005)	Literature study	Reviews more recent papers (since Le, Gibson, and Oxley 2005) which used modern econometric tools to analyse & understand how stock of HC is measured.	More detailed analysis of each research paper could supplement the study.	Presents research on modern econometric tools to measure HC.	Valuable research been done to improve on measurements of HC, particularly as a LV.
The estimation of HC in structural equations with flexible specification Lovaglio & Folloni, 2011 [Working paper]	Focus: statistical models for estimating Human Capital (HC) at disaggregated level (worker, household, graduates); New model proposed applied to measure causal impact of formal HC, on initial earnings for University of Milan (Italy) graduates.	New model, within Component Analysis framework, which allows for concomitant indicators in a structural model. Microeconomic (individual level) application.	3 data sources: Administrative Archive of Milan University (2001 to 2007); Employment Centres of the Province of Milan (2000 to 2007); & Italian Internal Revenue Service Archive (2000 to 2006). Estimates HC at bidimensional (EduHC & JobHC). Variables (formative, reflective & concomitant) selected based on economic theory. X1: formative for EduHC X2: formative for JobHC e1: Years of schooling J1: Evolution of the career path (observed period) m1: Legal duration Time / Time to obtain degree J2: Saturation of the career path (observed period) a1: Expected Graduation Age /Graduation age J3: Time (from degree) to obtain the Ist permanent contract a2: Graduation score/Mean graduation score. Y: endogenous indicators T: concomitant indicators y1: Income level (last available year) t1: Origin's household economic status y2: Annual income growth rate (2003-06) t2: Size of the origin's household	Limitations: Total costs of investing in HC, eg quality of education and other dimensions eg such as intelligence, ability, etc) difficult measure/data; Returns to reflective indicators 'hidden' in firm's outcomes and not measured by earnings; Other general shortfalls endogeneity and causality; Model does not allow for concomitant indicators to have impact on both endogenous indicators (eg income) as well as formative indicators.	Proposed estimation method simple yet versatile to fit complex relationship; When applying to HC, model proposed accounts for economic theory.	EduHC is largely attributed to years of formal schooling (e1) & to age of graduation (a1) & to lesser extent to graduation score (a2). Both HC dimensions significantly effect income levels variability (y1) & on earnings longitudinal dynamics (y2). Concomitant indicators have no direct significant impact on Y. Parents' economic status (t1) not have a direct impact on the workers' economic performance, however, it could be indirect but cannot be studied given model limitations. Need further research to extend model to allow for causal path between concomitant indicators and both endogenous & formative indicators
Structural Equation Models in a Redundancy Analysis Framework with Covariates. Lovaglio & Vittadini, 2014	Propose new method Generalized Redundancy Analysis (GRA) to fit SEM to measuring LV (eg HC). GRA allows a variety of relationships among composites,	GRA – Appendix A in the paper has the method specifications. Considering HC as a LV that must be estimated as composite, GRA method within structural model used. Supports the specification	3 data sources: Administrative Archive of Milan University (2001 to 20 07); Employment Centres of the Province of Milan (2000 to 2007); & Italian Internal Revenue Service Archive (2000 to 2006). Estimates HC at bidimensional (EduHC & JobHC). Variables:	GRA some specification limitations compared with PLS which require extensions. Also descriptive without inferential aspects.	GRA better estimates & specifies concomitant indicators in a model compared to ERA & PLS.	EduHC largely attributed to years of formal schooling & to age at graduation, which show the largest weights & to lesser extent graduation score. Household economic status concomitant indicator significant weight on EduHC, with larger contribution than graduation score, however does not have a

	endogenous variables, & external covariates to be specified & fit as shortfall of Lovaglio and Folloni (2011) research. New method applied to measure individual HC & impact on initial graduate (university of Milan) earnings.	of relationships consistently with a realistic economic process of HC accumulation.	X1: Formative for EduHC (X1 →f1) X2: Formative for Employ (X2 →f2) g1: Years of schooling s1: Employment security (career path) m1: Legal duration time to obtain degree s2: Saturation of the career path (observed career path) p1: Expected graduation age/graduation age s3: Time from degree to obtain the 1st contract p2: Graduation score/mean graduation score Y: Endogenous indicators; T: Concomitant indicators (T→f1, Y) y1: Income level (last available year) t1: Origin's household economic status y2: Annual income growth rate (2003–2006) t2: Size of the origin's household	Requires very detailed individual data often not available on a HH level. Only looks at graduates.		significant impact on earnings. Conclude much of effect of family background on HC is indirect & accounted for in the education choices & achievements. This is important because therefore only need to account for family background impact on HC construct & not its relationship on the reflective indicator (income). All selected formative indicators of JobHC contribute equally to score, however time from degree to first contract has limited significance.
Human capital estimation in Higher Education. Lovaglio, Vacca, and Verzillo 2016.	Specify & fit EM that represents mechanism of HC accumulatio n during higher education aimed at improving economic performance. Similar to Lovaglio & Vittadini (2014) study. Aim is to account for concomitant /external indicators which may have causal impact on both economic results (reflective indicator eg income earned) &/or Educational HC (the latent construct). Which was identified as a shortfall in the model specified in	GRA. Same as Lovaglio & Vittadini (2014).	Data sources: 1)data from University Administrative archives of four Universities of the Milan area, who graduated in 2007. 2) Archive of Employment Centers, Lombardy area. 3) Italian National Revenue Agency archive. 4) institutional databases (Ministry of Education, University & Research-MIUR-, Italian Committee for the Evaluation of Research -CIVR-) & web sources (ISI web of Knowledge). Variables: concomitant indicators (*t*1) reflecting parents' economic status *Formative indicators* *x*1 Expected graduation age/graduation age *s*1 Number of days from graduation to the first occupation *x*2 Graduation mark-cohort graduation mark *s*2 Saturation of the observed career *x*3 Condition at graduation *s*3 Saturation variation 2008–2009	Limitations: Selection bias as only takes into account graduates who are employed. Restricted data & short timeframe. Very detailed data needed on individual graduates to replicate in another region.	HC not simply measured as linear combination of pre-specified formative indicators, but are linear combinations that have some significant effect on the specified endogenous variables, thus allowing to select the best subset of pre-specified formative indicators in a predictive framework.	Results evidence that being employed prior to graduation positively affects labour market outcomes after graduation & however students with higher academic ability in times of economic depression postpone first entrance in labour market, thus avoiding lower wages & investing in additional (postgraduate) education.

	Lovaglio and Folloni (2011).		*x*4 Average number of citations per researcher in the Faculty in 2007 *x*5 Average impact per researcher in the faculty *x*6 Number of projects financed by MIUR over the period 2001–2003 in the faculty Reflective Indicators for *f*1 & *f*2 *x*7 Number of ISI publications per researcher in the faculty in 2007 *y*1 Last declared gross annual income *x*8 Faculty percentage of foreign students in 2007 *y*2 Average daily income (self-employed) *y*3 Average daily income (subordinate work)			
A least squared approach to latent variables extraction in formative-reflective models. Fattore, Pelagatti, and Vittadini 2018	A new least-squares based procedure for the extraction of LVs SEM with formative–reflective scheme developed & illustrated.	New least-square approach to measuring LV. Compares SEM, PLS-PM & new least-square approach using example of well known dataset on agricultural inequality, industrial development, political inequality for 47 countries. Authors freely provide software for R statistical package.	Applied to an example of agricultural inequality & industrial development (being LVs) & causes of political instability (LV). Formative indicators. Variables: Inequality of land distribution (Agricultural inequality); Percentage of farmers that own half of the land (Agricultural inequality); Percentage of farmers that rent all their land (Agricultural inequality); Gross national product per capita (Industrial development); Percentage of labour force employed in agriculture (Industrial development); Instability of executive (1945–1961) (Political instability); Number of violent internal war incidents (1941–1961) (Political instability); and Number of people killed as a result of civic group violence (1950–1962) (Political instability) for 47 countries.	Room for extensions which need further research but still already provides a valuable tool for managing formative–reflective schemes with clear advantages over PLS-PM & SEM.	Method suggested valuable alternative to PLS-PM & SEM since it is fully consistent with the causal structure of formative–reflective schemes & it extracts the factor scores without substantial identification or indeterminacy problems. Application of method straightforward.	Proposed procedure provides a valuable tool for managing formative–reflective schemes, such as HC, making the use of these important models more robust & effective (than SEM & PLS-PM) from the methodological & applied point of view.

CHAPTER 3

3 MEASURING HUMAN CAPITAL DISTRIBUTION IN SOUTH AFRICA USING A LATENT-VARIABLE APPROACH[35]

Abstract

The widely accepted definition of human capital (HC) recognizes HC as "the knowledge, skills, competencies and attributes embodied in individuals that facilitate the creation of personal, social and economic well-being" (OECD, 2001:18). HC is thus a complex concept to measure. Education is a fundamental dimension of HC and the education-based approach is the one most widely used to estimate HC in South Africa. However, the international literature recommends a latent-variable approach to measuring HC in a Structural Equation Modelling (SEM) framework. This study thus aims to measure HC in South Africa using a latent-variable approach and the National Income Dynamic Study (NIDS) wave 5 dataset. Thereafter, the profile of HC by socioeconomic status using the latent-variable and education-based approaches are presented. The reflective indicators included in the measurement model are health, parental education, educational attainment and schooling quality, selected as consistent with the South African context, human capital, and social reproduction theory. The findings are that the estimated latent HC, given the incorporation of multiple indicators, reveals more significant differences in HC levels, which the education-based proxy approach does not capture. The findings thus add value to the analysis of HC in the South African context of exceptionally high levels of socioeconomic inequality.

Keywords: human capital, education, inequality, latent construct

[35] The article is under review in the South African Journal of Economics. The co-author is Dr Fabio Correa.

3.1 Introduction

Human capital (HC) has long been recognized as invaluable to national wealth. Alfred Marshall (1920:564) states: "the most valuable of all capital is that invested in human beings". This translates to the HC of each household (HH) or individual being the most valuable national asset at the microeconomic level. Thinking about the concept of HC dates back as far as the seventeenth century, when Sir William Petty (1623-1687) was one of the first economists to try to define and measure it (Folloni and Vittadini, 2010). Before the twentieth century, research regarding HC was centered largely on enhancing national wealth accumulation[36] (Folloni and Vittadini, 2010). A shift in thought occurred with the introduction of microeconomic theories at the University of Chicago that regarded HC as a rational investment choice (Schultz, 1959, Becker, 1962, and Mincer, 1958). These authors were thereafter considered the 'founders' of HC theory. The 'new' theories were concerned with the "activities that influence future real income through the embedding of resources in people" – that is: through investments in HC (Becker, 1962:9). However, the definition of what HC is and therefore how it should be measured has continued to be refined. Laroche *et al.* (1999, in Le *et al.*, 2005) added individuals' innate abilities to the definition. The OECD (2001:18) reflected these advances, stating HC to be "the knowledge, skills, competencies and attributes embodied in individuals that facilitate the creation of personal, social and economic well-being". Thus, HC is today viewed as multifaceted and not directly observed (Folloni and Vittadini, 2010:267).

Le *et al.* (2005) identify three broad approaches to measuring HC. These are: an income-based measure, originally modeled by Farr (1853); a cost-based measure (Kendrick, 1976 and Eisner, 1989); and, lastly, various education-based methods (Barro and Lee, 1993 and Woßmann, 2003). Liu and Fraumeni (2021) categorised the HC measures into two broad approaches namely indicator-based and monetary measures. Towards providing researchers with comparable HC measures, Liu and Fraumeni (2021) discussed the methods and findings of six HC measures each of

[36]Controversy shrouded earlier discussions regarding HC as it was believed morally wrong to value humans as resources contributing to national wealth (Schultz, 1959 and Folloni and Vittadini, 2010).

which are estimated across at least 130 countries. The monetary approaches discussed by Liu and Fraumeni (2021) are those developed to help government planning by including HC and nonhuman capital in a framework of national wealth accounting. The World Bank's report titled "Changing Wealth of Nations 2018: Building a Sustainable Future (CWON)" included a monetary value for HC for 141 countries over the period 1995 – 2014 by applying the Jorgenson-Fraumeni approach (Liu and Fraumeni, 2021). A second HC monetary measure is that presented in the biennial Inclusive Wealth Report (IWR) which included HC monetary values for 140 countries over the period 1990 – 2014. The HC measure was estimated by following a framework developed by Klenow and Rodriguez-Clare (1997) and Arrow *et al.* (2012, 2013).

Indicator-based measures discussed by Liu and Fraumeni (2021) include four largescale composite indexes for HC. The indexes include the World Bank's Human Capital Index (HCI), United Nation's Human Development Index (HDI)[37], the Institute for Health Metrics and Evaluation Human Capital Index (IHME HCI)[38] and the World Economic Forum's Global Human Capital Index[39] (WEF GHCI). The indexes use various macroeconomic education and health indicators as the two basic dimensions of HC towards estimating a counties' distance from optimal HC (the distance from maximum HC attainable)

Since the late twentieth century, an additional approach to measuring HC has evolved namely a latent-variable approach. The approach integrated the indicator- and income-based approaches by extending the measured latent HC into monetary terms (Dagum and Slottje, 2000). This approach was suggested because HC is statistically defined as a multidimensional non-observed latent variable. To account for the multidimensional nature of HC, it is suggested the variables considered

[37] The HDI was developed to measure human development as a composite index and included in the first Human Development Report written by a team at the United Nations Development Programme (UNDP) led by Mahbub ul Haq (UNDP, 1990). South Africa ranks 114th out of 189 countries (UNDP, 2020)

[38] The IHME HCI measures the "expected years lived from age 20 to 64 years, adjusted for educational attainment, learning, and functional health status". South Africa ranks 144th out of 195 countries.

[39] The WEF GHCI has been constructed since 2015 with the updated methodologies and most recent estimates presented for 130 economics for 2017 in The Global Human Capital Report 2017 (World Economic Forum, 2017). South Africa ranks 87th out of 130 countries.

should broadly fall under the headings of education, work experience, indicators linked to career progression, health and, finally, income as an outcome of HC (Dagum and Slottje, 2000; Grossman, 2000; Dagum *et al.*, 2003; Vittadini *et al.*, 2003; Dagum *et al.*, 2007; Vittadini and Lovaglio, 2007; and Lovaglio, 2008, 2010). Given the intergenerational transfer of capital and welfare (Lovaglio, 2010), sociodemographic indicators should also be considered, for example parental socioeconomic status (SES). The literature on measuring HC as a latent variable has thus recognized the Structural Equation Modelling (SEM) framework as appropriate for measuring HC (Folloni and Vittadini, 2010 and Laverede-Rojas *et al.*, 2019).

Most recently, the World Bank (2018a) developed a Human Capital Index (HCI) to measure, at a country level, the HC which "a child born today can expect to attain by her 18th birthday, given the risks of poor health and poor education where she lives" (World Bank, 2018a:3). The purpose of the index was to provide policymakers with an understanding of how much development and growth is being lost in existing HC gaps (World Bank, 2018). Consequently, the Human Capital Project (HCP) aimed to work with countries to ensure better investments into HC. The South African HCI estimate was 0.41 (out of a possible 1) and thus ranked 126 out of 157 countries (World Bank, 2018). This confirmed the importance of further research on understanding the multidimensional nature of HC in South Africa.

In the next section, literature on the importance of HC as a determinant of socioeconomic inequality in South Africa is discussed, followed by an examination of the approaches used previously to measure HC in South Africa. Thereafter the methodology, including the data, analytical approach and variables used in this paper, are explained. In section 3.5 the results, including the distribution of the latent HC estimate, on an individual level and by different SES indicators, are presented. The final section draws conclusions.

3.2 Human Capital as a determinant of inequality in South Africa

South Africa has the highest level of wealth inequality and one of the world's highest levels of income inequality (Orthofer, 2017). The gini coefficients for wealth and income were estimated as 0.95 in 2017 (Orthofer, 2017) and 0.63 in 2014 (World Bank, 2019), respectively, without significant decreases in the inequality estimates over the 26 years since the end of Apartheid (Chatterjee *et al.*, 2020). Since the end of the apartheid most of the research on income distribution in South Africa focused on inter-racial income distribution (Van der Berg and Louw, 2003). Van der Berg (2014) found that inequality amongst racial groups has increased. Several factors have contributed to this, most importantly the unequal distribution of skills and inequality in educational outcomes, which Spaull (2015) highlighted as driven by inequalities in schooling quality. Furthermore, findings indicate that income inequality in South Africa is generated in the labour market (Leibbrandt *et al.*, 2012; Van der Berg, 2014; and Hundenborn *et al.*, 2016).

The impact of HC in South Africa on poverty and the labour market has been researched by authors including Lam (1999), Liebbrandt *et al.* (2010), Van der Berg (2010), Pellicer and Ranchhod (2012), Branson *et al.*, 2013 and Wanka (2014). Lam (1999) examined educational inequality and its impact on earnings in South Africa and Brazil. Included is the intergenerational transmission of educational inequality. Lam (1999) found that, firstly, schooling plays a large role in explaining earnings inequality in both countries and, secondly, educational inequality strongly influences intergenerational transmission.

Liebbrandt *et al.* (2010) noted the importance of education and health policies in South Africa in equipping the labour force, particularly those disadvantaged by apartheid, with the skills required for active participation in a productive workforce and so in earning the income necessary to address poverty and inequality. This is aligned with HC theory regarding investments in HC, including education, on-the-job training and health, which are rational choices towards increasing earnings. Van der Berg (2010) furthermore showed that, rather than access to political power or wealth, wage inequality was a major driver of income inequality in South Africa and was itself the result of unequal HC distribution (measured as the educational attainment and the quality thereof). He thus concluded that improved education quality was required both for future economic growth and to reduce poverty and income inequality.

In studying HC accumulation and inequality traps in South Africa, through assessing theoretical frameworks relating inequality and decisions to invest in post-secondary education, Pellicer and Ranchhod (2012), concluded that government policies are needed to provide quality education that will lead to accumulating the necessary HC to reduce inequality in future. Pellicer and Ranchhod (2012) furthermore highlighted that an inequality trap is evident in South Africa, where high income inequality leads to low levels of skill accumulation, perpetuating income inequality. This provides grounds for an additional conceptual perspective on the relationship between education and earnings, namely the theory of social reproduction (Chiappa and Mejias, 2019).

In the context of HC, the social reproduction perspective proposes that the social stratification order is reinforced by the education system (Bourdieu and Passeron, 1977). The idea of social reproduction stems from Marx's statement that "every social process of production is, at the same time, a process of reproduction" (Marx, 1977:711). Social reproduction theory thus highlights the intergenerational transmission of SES, given the accumulated disadvantages accrued during childhood by being born into a family with a lower SES and parents with low levels of education (Breen and Johnson, 2005). The OECD (2014) indicates that the combined impacts of social background and HC effects reproduce SES and thus a circular relationship between SES and HC is perpetuated. It is this reproduction of social order, highlighted by Bowles and Gintis (1975), that is missing from microeconomic HC theory. This furthermore highlights the potential for endogeneity when studying HC and SES. This is particularly relevant in South Africa. Spaull (2015), for example, states that a child's future socioeconomic outcomes (for example: employment, income and education) in South Africa can be predicted with a certain level of accuracy by the age of seven. Children from poorer homes do not acquire basic skills, which reflects their future productivity, and ensures they remain perpetually disadvantaged. Wealthy individuals, HHs and nations are expected to spend more on education and consequently acquire more HC (Chani *et al.*, 2012). This suggests that an unequal distribution of income/wealth may negatively impact HC inequality and vice-versa (Chani *et al.*, 2012). This would furthermore explain why many talented South African citizens are unfortunately not likely to reach their full potential, which negatively impacts much needed social and economic development (Keeton, 2014).

3.3 An alternative approach to measuring HC

Despite education being a fundamental element of HC, Le *et al.* (2005) note that education measures are simply a proxy for HC and do not measure the multidimensional aspects of HC directly (United Nations, 2016). When critiquing the education-based method, Le *et al.* (2005) highlighted that the different proxies originally used at a national level were: adult literacy rates, school enrollment ratios, and average years of schooling. However, these education proxies were noted as misspecifying HC, specifically on a theoretical basis (Wößmann, 2003) and fail to account for the multidimensional nature of HC (World Bank, 2018). The most used proxy, namely average years of schooling[40], was critiqued for not allowing for returns to education and the quality of the education received. For education to make a meaningful contribution to HC and a country's economic growth and development, it needs to be of high quality and provide the skills demanded in the economy. Given the almost non-comparable differences in education quality in South Africa (Spaull, 2019), there is sound motivation for moving beyond the education-based method of measuring HC. As a proxy for HC, education attainment is thus subject to proxy measurement error, which Wendelspiess Chávez Juárez (2015:119) defined "as a deviation of the observed variable with respect to the underlying concept we actually want to measure".

Given that HC cannot be observed directly, it has been statistically defined as a multidimensional composite variable and thus needs to be measured in a latent

[40] The national stock of HC used in production would thus be measured as the highest educational attainment of workers (Wöβmann, 2003).

variable framework (Folloni and Vittadini, 2010 and Laverede-Rojas *et al.*, 2019). Dagum and Slottje (2000) described household HC as the "multidimensional nonobservable[41] construct generated by personal ability, home and social environments, and investments in the education of the HH head and spouse whose effects are indirectly measurable by means of the present value of a flow of earned income throughout an individual life span" (Dagum *et al.*, 2007: 581). Dagum and Slottje (2000) were among the first to formalize the latent-variable approach to measuring HC in a breakthrough study that introduced an actuarial mathematical approach to convert the composite construct HC to monetary values. This was then used to calculate the distribution of HH HC in the United States of America (USA). This method was repeated by Dagum *et al.* (2003) who studied the relationships between wealth, debt and HC. Authors including Dagum *et al.* (2007), Vittadini and Lovaglio (2007) and Lovaglio (2008) extended the method developed by Dagum and Slottje (2000) to ensure estimates are consistent with the economic definition of HC suggested by Dagum *et al.* (2007: 581). The authors, studying HH HC in the Italian and USA context, concluded that HC should be estimated in a formative-reflective model[42] in a SEM framework.

Vittadini *et al.* (2003), noted, that there are numerous ways in which a latent variable can be defined, each of which will have implications for the statistical model used. For example, HC may be estimated by means of Factor Analysis if the latent HC is identified as a factor representing the variance in the observed reflective indicators. Ultimately the model specified should reconcile with the meaning and definition of HC applied in the study and empirical evidence (Coltman *et al.*, 2008).

Laverede-Rojas *et al.* (2019) analyse the benefits of measuring HC as a latent variable in a SEM framework in comparison to an education-based approach, by comparing their predictive power for economic growth. The study spans from 1970 to 2011 and includes 91 countries with different development levels. They find that

[41] Dagum and Slottje (2000) used the following variables as indicators of personal ability, home and social environments: age of HH head, years of schooling of HH head and spouse, years of fulltime work of HH head and spouse, marital status, gender of HH spouse, region of residence, number of children, total wealth and total debt.

[42] A formative-reflective model suggests that HC is simultaneously obtained from both formative and reflective indicators of the latent construct.

years of education has reduced predictive power since the 1990s, which they attribute to schooling having become homogenized, with almost universal enrolment. This could be an answer to the question raised by Castelló-Clement and Domenech (2017) who investigated why the reduction in education attainment inequality over the years, has not been accompanied by falling income inequality.

It is therefore evident that policymakers would benefit from applying a multidimensional measure of HC to help guide policies to ensure socioeconomic development. The variables used to measure latent HC have varied slightly between studies. When selecting variables to estimate HC at a disaggregated level in South Africa it is important to consider data availability, economic definition and context. From previous studies (Dagum *et al.,* 2007; Vittadini and Lovaglio, 2007 and Lovaglio 2008/2010), the variables in the model should be consistent with the definition of HH/individual HC and fall into the following broad areas: education indicators, demographic indicators, on the job training and participation, adult skills and family characteristics. These variables are consistent with HC and social reproduction theory. Income is an additional variable which should be considered in accordance with the income-based approach (Vittadini *et al.*, 2003).

This study thus estimates HC on a microeconomic level as a latent construct, in a SEM framework, and uses the results to analyse socioeconomic inequalities. The meaning of HC as defined by the OECD (2001) was used to develop the conceptual path model in this study and HC is thus defined as a non-observable construct, developed in home and social environments that best determine socioeconomic wellbeing. Furthermore, the South African context, data availability and variables suggested by Dagum *et al.* (2007), Vittadini and Lovaglio (2007) and Lovaglio (2008/2010) guided the development of the conceptual model.

3.4 Method

3.4.1 Data and Sample

Primary data from the National Income Dynamics Study (NIDS) dataset was employed. The NIDS is a panel dataset compiled biennially by the South African Labour and Development Research Unit (SALDRU) at the University of Cape Town (UCT). The first wave of data was released in 2008 (NIDS, 2018). The initial wave was administered to a nationally representative sample of approximately 28 000 individuals from 7 300 HHs, but grew with each wave as HHs expanded to approximately 37 000 individuals[43] and 13 719 HHs in wave 5. The most recent, Wave 5, conducted in 2017 and released in 2018, is used in this study (Brophy *et al.*, 2018). The two main reasons for using the NIDS dataset were, firstly, the comprehensive nature of the socioeconomic data collected[44], and, secondly, that this research forms part of a larger study, which aims to analyse HC using different measurement approaches and the impact of HC on inequality. Using the same dataset makes the different HC measurements comparable. The sample was restricted to employed individuals in order to analyse the distribution of HC in the labour market (the market in which HC is employed and income inequality in South Africa is generated (Leibbrandt *et al.*, 2012)). Employed individuals without missing observations for the variables in the model were included in the sample and thus restricted to 2 771 individuals. Post-stratification weights were applied[45] to allow for population estimates (SALDRU, 2017).

3.4.2 Analytical approach

This study employs a SEM framework, a multivariate statistical analysis method, which combines factor analysis and multiple regression (Bollen, 2002). The analytical approach includes three steps: 1) Exploratory data analysis; 2)

[43] The total number of individuals successfully interviewed in Wave 1 and Wave 5 were 26 776 and 37 368 respectively (Brophy *et al.*, 2018).
[44]Examples of variables which will be used from the NIDS dataset include school administration data, grade repetition, school quintile, literacy, educational attainment, employment history, income, parents' education, stunting, mortality history, assets, debt and health.
[45] Weight variable w5_wgt (post-stratified weight for Wave 5) was applied.

Confirmatory Factor Analysis (Measurement model); and 3) Analysis of the profile of HC by SES indicators, estimated using a latent-variable (HC_L) and education-based approach (HC_E). The first two steps were conducted to establish the model and estimate HC_L per person in the sample. A sensitivity analysis is included to confirm the robustness of the results. Step 3 was included to explore HC distribution using the education-based and latent-variable approaches; and the HC profile by SES indicators, namely, income, gender, race, and geographical area.

3.4.2.1 Conceptual measurement model and variables

Different path models were explored to estimate the latent variable HC[46] within a SEM framework and it was identified that the latent variable HC is best estimated in South Africa by applying a reflective[47] Confirmatory Factor Analysis (CFA) measurement model[48]. Jarvis (2003) identified that a reflective model is motivated if the indicators all represent the underlying construct, thus potential proxies of HC, and are correlated. Please see the correlation matrix in Appendix 3.A. Furthermore, the indicators should be interchangeable, and dropping one indicator should not alter the conceptual meaning of the latent construct (Jarvis, 2003). The reflective indicators selected in the measurement model of HC are parental education, education attainment, school quintile and self-perceived health status (see Equation 1). The choice of reflective indicators is consistent within the South African context (Spaull, 2015) and social reproduction and human capital theories.

---

[46] A formative-reflective model like those developed in the Italian and USA contexts by Vittadini *et al.* (2003); Dagum *et al.* (2007); Vittadini and Lovaglio (2007); Lovaglio (2008; 2010) and Folloni and Vittadini (2010) was explored. However, using the NIDS data and the statistical programs R.4.0.3 (packages *lavaan*, *ltm* and *psych*) and STATA (*sem* package) revealed that the formative-reflective model does not converge to provide statistically significant results: the coefficients of the formative indicators were all 0 and *p*-values over 0.9. These results thus indicate that the variance between the observed indicators does not sufficiently converge to estimate HC using a formative-reflective model and the latent HC variable is not identifiable. Thus, the model is not appropriate in the South African context using the NIDS dataset. The results of the models explored were confirmed by the Statistical Consulting Service at UCT.

[47] Reflective CFA models attempt to account for the variance and covariance of the observed variables (Jarvis *et al.*, 2003).

[48] Coltman *et al.* (2008) noted that proposed CFA measurement models should be specified using theory and empirical evidence to confirm and estimate latent variables (Coltman *et al.*, 2008). Thus, both reflective and formative CFA models may be validated, which means that the observed indicators can be specified as causal in nature (here HC would be estimated in a formative model specifying the indicators as causes of HC) or reflective in nature (here HC would be estimated in a reflective model specifying that the observed indicators are representatives of HC).

$$HC_{Li} = f\left(edu_i, fthedu_i\,, mthedu_{i,} schoolquin_{i,} health_i\right) + \varepsilon_i \qquad (1)$$

Where HC_{Li} is the estimated latent variables HC; edu_i the education; $, fthedu_i$ the father's education; $mthedu_{i,}$ the mother's education; $schoolquin_{i,}$ the school quintile; $health_i$ health; and ε_i the error term, associated with each specified variable, of the ith observation. The measurement model can be written as a set of five regression equations, as captured in Equations 2 to 6, or illustrated in the conceptual measurement model presented in Figure 3.1. The motivation for the indicators in the model is provided below and the variables are detailed in Table 3.1. It should be noted that Figure 3.1 does not imply causality, but estimates how well the specified observed indicators represent the HC_L (Bagozzi, 2011) Following the goodness of fit analysis and confirmation of the measurement model, HC_L is predicted for further analysis (Coltman *et al.*, 2008 and StataCorp, 2013).

$$edu_i = \alpha_i + \beta_i HC_L + \varepsilon_i \qquad (2)$$

$$fthedu_i = \alpha_i + \beta_i HC_{Li} + \varepsilon_i \qquad (3)$$

$$mthedu_i = \alpha_i + \beta_i HC_{Li} + \varepsilon_i \qquad (4)$$

$$schoolquin_i = \alpha_i + \beta_i HC_{Li} + \varepsilon_i \qquad (5)$$

$$health_i = \alpha_i + \beta_i HC_{Li} + \varepsilon_i \qquad (6)$$

Where α_i is the intercept; β_i the path coefficient; and ε_i the error term associated with the specified observed indicator e.g. health. The path diagram is illustrated in Figure 3.1.

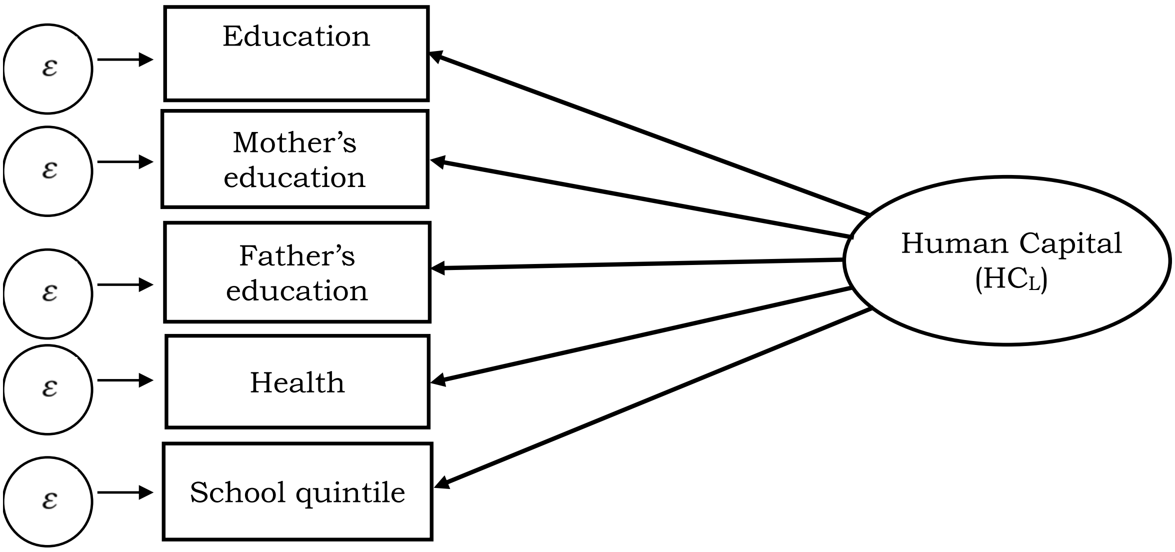


Figure 3.1: Path diagram: Reflective measurement model of HC.

Source: Authors' construction

Equations 2 to 6 are used to estimate the path coefficients between the specified unobserved latent (HC_L) and observed indicators in the data and thereafter $\widehat{HC}_L$ per observation are predicted for further analysis. The observed indicators are consistent with the OECD (2001) definition of HC, empirical evidence in South Africa (Van der Berg, 2010; Pellicer and Ranchod, 2012 and Spaull, 2015 & 2019), HC and social reproduction theory, as well as the variables suggested by authors aiming to account for the multidimensional nature of HC internationally (Dagum and Slottje, 2000; Grossman, 2000; Dagum *et al.*, 2003; Vittadini *et al.*, 2003; Dagum *et al.*, 2007; Vittadini and Lovaglio, 2007; Lovaglio, 2008, 2010; and Kraay, 2018)[49]. The measurement of each variable is explained in Table 3.1.

[49] The authors highlight the variables considered should broadly fall under the headings of education, work experience, indicators linked to career progression, health and, lastly, income as an outcome of HC.

Table 3.1 Measurement of the variables included in the analysis.

Variables representing the reflective indicators for estimating HC_L.	*Variables employed in the analysis of the distribution of HC*
Education (Categorical)	Gender (Dummy)
1. Primary or less; 2. Incomplete secondary; 3. Matric; 4. Some tertiary; 5. Bachelor or higher	1. Male; 0. Other
Self-perceived health status (Categorical)	Race (Categorical)
1. Poor; 2. Fair; 3. Good; 4. Very good; 5. Excellent	1. African; 2. Coloured; 3. Asian/Indian; 4. White
School quintile (Categorical)	Geographical area (Dummy)
1. No-fee school; 2. No-fee school; 3. No-fee school; 4. Low-fee school; 5. Fee charging schools (private & public)	1. Urban; 0. Other
Mother's Schooling (Categorical)	Individual monthly labour income
1. Primary or less; 2. Secondary; 3. Matric; 4. Tertiary	Continuous variable (Min: 16; Max: 189 000)
Father's Schooling (Categorical)	Variable for estimating HC_E
1.Primary or less; 2. Secondary; 3. Matric; 4. Tertiary	Highest educational attainment (Min: 0; Max: 18)*

Notes: *The NIDS education attainment data was coded as per Branson and Leibbrandt (2013), with the exception of masters and doctoral degrees which was coded as 18 years.

The inclusion of parental education is supported by social reproduction theory and the empirical evidence of the intergenerational transmission of socioeconomic outcomes in South Africa (Pellicer and Ranchod, 2012). Parental education furthermore represents the intergenerational transfer of cognitive and non-cognitive skills and health outcomes (Lundborg *et al.*, 2018). Employing parental education allows for social reproduction theory in the production of HC to be tested. School quintile is included because empirical evidence suggests that in South Africa the school attended plays an important role in educational attainment. This is because high-fee-paying schools are associated with better education quality (HC theory), as well as the reproduction of education outcomes where only the wealthy are able to afford to send their children to better functioning fee-paying schools (social reproduction theory).

Education attainment and health are included in accordance with HC theory which specifies these as investments in HC (Becker, 1964) and facilitating social mobility. Becker (1964:3) included health in the definition of HC, stating that HC is the "knowledge, information, ideas, skills, and health of individuals". Another example of previous work which included health as a component of HC is the World Bank's

HCI, which included adult survival and under-5 stunting rates as proxies for health (Kraay 2018).

Empirical evidence from South Africa has found that higher education attainment and better health have positive impacts on socioeconomic wellbeing (van Broekhuizen; 2011; Branson and Leibbrandt 2013 and Nwosu, 2015). Self-reported health status is used to proxy actual health. Nwosu (2015), identified the arguments for and against using subjective self-perceived health status variables. The arguments in support are the ability to capture various dimensions of health (Bound, 1991) and future mortality (de Salvo *et al.*, 2005 and Ardington and Gasealahwe, 2012). However, nonrandom measurement errors of self-perceived health indicators highlight potential bias (Kreider, 1999). The conclusion reached by Nwosu (2015) was that, in capturing health and labour market participation relationships, self-perceived health responses may be more effective than reported individual illness or symptom responses (Bound, 1991). Furthermore, Ardington and Gasealahwe (2012) stated that, despite some measurement error, the health variables in the NIDS dataset are useful for the analysis of health and SES relationships.

The results from the reflective model will highlight potential areas requiring additional investments to reduce HC and productivity gaps in South Africa.

3.4.2.2 Exploratory Data Analysis

The number of observations, mean and standard deviation of the variables relevant to the model and analysis of the distribution of HC_L are presented in Table 3.2. Included in the exploratory data analysis, a log-linear earnings function, specified in Equation 7, is estimated to explore the associations between the observed HC reflective and socioeconomic indicators and labour income, as per microeconomic HC theory (Schultz, 1959, Becker, 1962 and Mincer, 1958). The correlation coefficients are presented in Table 3.2.

The log-linear earnings function is:

$$\log Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots \beta_1 X_{i1} + \beta_k X_{ik} + \varepsilon_{i,} \qquad i = 1, 2, \ldots\ldots n. \qquad (7)$$

Where $\log Y_i$ is the log of net monthly labour income per capita in ZAR for observation i and β_0; $X_1, X_2, \ldots\ldots X_{ik}$ are the observed study variables (presented in Table 3.2); $\beta_1, \beta_2, \ldots \beta_k$ are the respective coefficients; ε_i is a vector of residuals per observation (error term); and n is the sample size. Equation 7 is consistent with log-linear earnings functions adopted by economists to estimate the private returns to investments in HC as well as differences in earnings as per socioeconomic indicators (van Broekhuizen, 2011 and Branson and Leibbrandt, 2013). The estimates are thus included in the descriptive statistics.

The data was tested to identify the suitability for structure detection and factor and SEM analysis. The Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy (MSA) test and Bartlett's sphericity test, were therefore performed. The KMO indicates the proportion of variance in the variables that might cause underlying factors, with a result above 0.6 verifying the sample to be acceptable (Pett *et al.*, 2003). Bartlett's sphericity test identifies whether there is significant variation between the survey items, hypothesizing that the correlation matrix is an identity matrix. Rejecting the null hypothesis would indicate that variables are related and thus suitable for factor and SEM analysis (Treiblmaier and Filzmoser, 2010) and the measurement model may be specified (Loehlin and Beaujean, 2017).

3.4.2.3 Confirmatory Factor Analysis (The measurement model)

CFA is a way of analysing specified measurement models (StataCorp, 2013) and the validity of the relationships between the observed indicators (manifestations) and proposed constructs (Loehlin and Beaujean, 2017). Thus, for the purpose of obtaining a reliable measure of HC_L a CFA measurement model was estimated as presented in the conceptual model above (Figure 3.1). The statistical significance of the relationships between the suggested indicators and the proposed HC_L were estimated in order to judge the reliability of the observed indicators. The goodness of fit of the model was assessed using the coefficient of determination (CD), comparative fit index (CFI), Tucker-Lewis index (TFI), standardized root mean squared residual (SRMR) and root mean square error of approximation (RMSEA). The indices were benchmarked as follows: CFI values ≥ .90 (Bentler, 1990); TFI values ≥ .90 (Hu and Bentler, 1999); SRMR values ≤ .06 (Hu and Bentler, 1999) and RMSEA values ≤ .08 (Browne and Cudek, 1993). Furthermore, a CD close to 1 (R^2) would indicate that the model fitted the data well. Following model confirmation, a continuous variable HC_L was predicted[50], which is factor score[51] extraction conditional on the value of the observed variables (StataCorp, 2013). The statistical programme STATA (*sem* package) was used. The path coefficients and goodness of fit results are presented in Table 3.3.

Furthermore, a sensitivity analysis was conducted, using two approaches to test whether the estimated measure of HC_L is robust. Firstly, the reflective CFA approach is retained but the specified model in Figure 3.1 is adjusted in two ways: 1) the health observed indicator is removed to predict the latent variable referred to as HC_h; and 2) the father's educational attainment observed indicator is removed to predict the latent variable referred to as HC_f. The results are compared by presenting goodness of fit results and standardized coefficients of the reflective indicators in Table 3.4.

Secondly, an alternative data reduction technique (DRT), namely Principal Component Analysis (PCA), was used to assess whether the results are sensitive to the estimation approach. PCA is a statistical method to reduce the number of variables into a few principal components, while maintaining as much of the variance as possible to reproduce the data structure for further analysis (Mooi *et al.*, 2018). Thus, PCA is an alternative approach to measuring a composite HC (referred to as HC_p) to confirm the findings on HC_L in South Africa. The first principal component (PC1), which accounts for the largest possible variance in the

[50] HC_L was predicted as a continuous variable per person in the sample using z-transformations and thus has a mean of 0. Negative and positive HC_L are expected.
[51] The factor scores are estimated by means of a linear regression, using the fitted model's mean vector and variance matrix.

dataset, is used to predict HC_p. The factor weights per component are presented in Table 3.5.

3.4.2.4 Analysing the profile of HC by SES indicators

Firstly, to map the concentration distribution of HC, Kernel density (k-density) curves[52] for HC_L, HC_h, HC_f, HC_p and HC_E are illustrated in Figure 3.2. Thereafter, a histogram is presented to compare the frequency distribution of HC_L and HC_E. Finally, Table 3.6 and Table 3.7 present the profile of HC_L and HC_E, respectively, by socioeconomic indicator. The analysis by socioeconomic indicator is motivated by the Sustainable Development Goals (SDGs) which highlight the importance of studying the outcomes of "the poorest, most vulnerable and those furthest behind" (United Nations, 2015:37). In South Africa, these indicators include race, gender, geographical region and labour income (StatsSA, 2019). The figures were constructed using the statistical program R.4.0.3.

3.5 Results and discussion

3.5.1 Descriptive Statistics

The descriptive statistics for the sample datasets, including the number of observations, mean (for continuous variables) and percentage distribution of observations per category (for the categorical and dummy variables), are presented in Table 3.2 below. Furthermore, the partial regression coefficients estimated from Equation 7 are presented.

[52] K-density curves illustrate the density of a variable from the observations of the variable.

Table 3.2: Sample data descriptive statistics and OLS coefficient estimations.

Variable	Observations and descriptive statistics (Post-stratified weights applied)	OLS regression model. R^2 = 0.4471	
Y (individual monthly labour income)	Obs 5 321 615; Mean 9 370.77; Std dev 14 745.83		
logY (log labour income per month)	Obs 5 321 615 Mean 8.434; Std dev 1.226		
Education attainment	5 321 615	Coefficient	Std. error
1 Primary or less	9.93%	Reference group	
2 Secondary	34.14%	0.30*	0.00
3 Matric	22.01%	0.67*	0.00
4Some tertiary	22.73%	1.32*	0.00
5 Bachelor degree or more	11.19%	1.98*	0.00
Self-perceived health status	5 321 615		
1 Poor	1.78%	Reference group	
2 Fair	6.23%	-0.06	0.74
3 Good	24.17%	0.09	0.55
4 Very good	32.81%	0.04	0.76
5 Excellent	35.01%	0.12	0.43
School quintile	5 321 615		
1 No-fee school	24.59%	Reference group	
2 No-fee school	20.26%	-0.01	0.93
3 No-fee school	26.16%	-0.03	0.72
4 Low-fee School	10.94%	0.11	0.43
5 Fee charging school	18.05%	0.16	0.23
Mothers education	5 321 615		
1 Primary or less	57.16%	Reference group	
2 Secondary	21.70%	0.07	0.43
3 Matric	11.11%	-0.04	0.82
4 Tertiary	10.03%	0.13	0.34
Fathers education	5 321 615		
1 Primary or less	54.22%	Reference group	
2 Secondary	23.31%	0.21**	0.012
3 Matric	11.61%	0.10	0.46
4Tertiary	10.87%	0.11	0.49
Gender	5 321 615		
0 Female	46.50%	Reference group	
1 Male	53.50%	0.49*	0.00
Race	5 321 615		
1 African	78.90%	Reference group	
2 Coloured	6.92%	0.18***	0.07
3 Asian/Indian	1.61%	0.57*	0.005
4 White	12.57%	0.60*	0.00
Geographical area	5 321 615		
0 Otherwise	25.60%	Reference group	
1 Urban	74.40%	0.10	0.16

Notes: Statistically significant (*p < 0.01; ** p<0.05; *** p<0.10).

Source: NIDS wave 5; authors' calculations. Post-stratified weights applied.

The results indicate that the majority of the sample have less than matric, have good health or better, attend no-fee schools, have parents with primary schooling or less, are females, African and reside in urban areas. It is estimated that educational attainment has improved over the generations, with over 50% of parents obtaining primary schooling or less, compared to 9.93% of the current generation. This is potentially a result of the reduction in variance in schooling curricula, educational attainment and almost universal schooling enrolment, which Laverde-Rojas *et al* (2019) referred to as 'educational homogenization'. However, higher levels of educational attainment (a Bachelor's degree or more) remain low, at 11.19% of employed individuals. This is concerning, given that education returns to earnings are high and increasing for degree graduates in South Africa (van Broekhuizen; 2011 and Branson and Leibbrandt; 2013) and income inequality in South Africa originates in the labour market. It is expected that without increases in higher levels of education attainment, income inequality will persist. The percentage of individuals with Bachelor's degree or more (7.65%) was even lower for the economically active population in the larger Wave 5 dataset.

The increased returns to higher levels of education is supported by the OLS regression results presented in Table 3.2. Having a Bachelor's degree or more is correlated with earnings 198% higher in comparison to those with primary schooling or less[53]. Only one of the parental education estimated coefficients, and none of the school quintile, geographical area and self-perceived health regression coefficients, were statistically significant. However, while controlling for all other model variables, the gender and race results were consistent with expectations, including: males are expected to earn on average 49% more than females[54]; and Whites, Asian/Indians and Coloureds are expected to earn 60%, 57% and 18% respectively more than Africans.

The results from Bartlett's sphericity and KMO tests confirmed sample data adequacy and sufficient variation between the proposed variables for SEM analysis. Bartlett's sphericity test was highly significant (χ^2 = 3204.67, ρ = 0.0000) and the KMO test coefficient was 0.739 and thus the factor analysis could be conducted.

[53] Controlling for all other variables in the model.

[54] Comparing males and female in the same jobs StatsSA (2019) found males earned on average 30% more than females. The finding of a wage gap of 49% is thus larger than the StatsSA (2019) finding and should be noted as the estimates based on the specifications of Equation 7 and the sample studied.

3.5.2 Confirmatory factor analysis (CFA) measurement model

The CFA measurement model had the objective of achieving a reliable measure of HC. The model was specified as standardized, thus the variance of HC_L is set to 1. All the reflective indicators are positive, statistically significant, reflective indicators of HC_L as presented in Table 3.3 below. The fathers' and mothers' education coefficient estimates are both above 0.8 and are thus important proxies for latent HC. The education attainment and school quintile indicators are both significantly associated with HC_L, with coefficients of 0.541 and 0.513 respectively. Health was included in the model and although it was estimated as having the lowest factor loading (0.158), its contribution to the HC_L was statistically significant. The health indicator was retained in the model, since HC theory suggests investments in health are representative of HC, as well as its positive impact on labour force participation (Nwosu, 2015), educational attainment and labour productivity (Suhrcke *et al.*, 2006). Nwosu (2015), using the NIDS dataset in a cross-sectional analysis, found statistically significant impacts of health (using different measures[55]) on labour force participation of between 20% and 33%. Table 3.3 also includes the gini coefficient[56] estimates of each variable, indicating health to be the least unequally distributed and father's education the most.

[55] The econometric techniques used included instrumental variables, censored quantile regression and Blinder-Oaxaca decomposition.

[56] The gini coefficient estimates the deviation from perfect equality and ranges between 0 and 1, with 0 indicating perfect equality and 1 perfect inequality.

Table 3.3: CFA measurement model: Standardized coefficient estimates, goodness of fit results and predicted HC estimates.

	Post-stratified weights applied					
	Effect	Gini index	Std. coefficient	S.E.	R^2	P-value
HC ~	Education	0.227**	0.541	0.024	0.30	0.000
	Mother's education	0.288	0.823	0.021	0.68	0.000
	Father's education	0.290	0.831	0.021	0.69	0.000
	Health	0.137	0.158	0.034	0.02	0.000
	Quintile	0.283	0.513	0.033	0.24	0.000
	Goodness of fit results*	Test	Benchmark	Result	Good fit	
		CD	Close to 1	0.837	Yes	
		SMRM	≤ .06	0.017	Yes	

Notes: *Applying post-stratified weights restricts the goodness of fit results to the size of residuals tests. **Estimated as per the five education brackets used in the estimate of HC_L.

Source: NIDS wave 5; authors' calculations. Post-stratified weights applied.

When applying the post-stratified weights, the goodness of fit tests were restricted to residuals. The CD value of 0.837 indicates that 83.7% of the variance in the observed indicators is captured in the HC_L construct. Providing a broader range of goodness of fit test results, the unweighted estimates, illustrated in Appendix 3.B, confirm the goodness of fit of the model and the estimated path-coefficients. Thus, social reproduction in HC development is identified. The equation level goodness of fit confirmed the importance of father's and mother's education, with R^2 values of 0.69 and 0.68, respectively. Thus 69% and 68% of the variance in father's education and mother's education is captured in the HC_L construct, respectively. Appendix 3.B below indicates the Chi-square goodness of fit estimate was 25.396 with a *p*-value of 0.000. The null hypothesis for the Chi-square test is that the model fits the observed data perfectly, thus rejecting the null-translates-to-poor model fit. However, it is noted that the Chi-square test is sensitive to sample size, with larger sample sizes decreasing the *p*-value where there may be only a trivial misfit (Babyak and Green, 2010). Therefore, it is recommended that chi-square test results be considered in combination with the CFI, TFI, SMRM, RMSEA and other fit tests (Alavi *et al.*, 2020).

The sensitivity analysis results presented in Table 3.4, highlight how removing the health indicator improved the goodness of fit, particularly evident when considering unweighted goodness of fit tests. However, the standardized coefficients show only negligible changes in comparison to the full model. Removing, father's education attainment illustrates good model fit (albeit a lower CD and greater SRMR) and the standardised coefficient estimates remain comparable with the HC_L estimates, with only negligible differences.

Table 3.4: Adjusted CFA measurement models: Standardized coefficient estimates and goodness of fit results.

		HC_h (no health)		HC_f (no father's educational attainment)	
	Effect	Estimate	Std Error	Estimate	Std Error
HC ~	Education	0.538	0.024	0.558	0.033
	Mother's education	0.821	0.022	0.795	0.034
	Fathers education	0.834	0.022		
	Health			0.183	0.036
	Quintile	0.513	0.033	0.523	0.034
Goodness of fit results	Benchmark	Estimate	Good fit	Estimate	Good fit
Post-stratified weights applied*					Yes
CD	Close to 1	0.837	Yes	0.721	Yes
SMRM	≤ .06	0.002	Yes	0.016	Yes
No weights applied**					
CFI	≥ .90	0.999	Yes	0.987	Yes
TFI	≥ .90	0.999	Yes	0.962	Yes
SMRM	≤ .06	0.007	Yes	0.021	Yes
RMSEA	≤ .08	0.011	Yes	0.053	Yes
CD	Close to 1	0.836	Yes	0.711	Yes

Notes: *Applying post-stratified weights restricts the goodness of fit results to the size of residuals tests. **The unweighted estimates of the CFA measurement model confirm the estimated path-coefficients and goodness of fit of the model.

Source: NIDS wave 5; authors' calculations. Post-stratified weights applied.

The factor scores presented below in Table 3.5, estimated using PCA, indicate comparable variance estimations identified using the CFA measurement model to predict HC_L, as indicated in Table 3.3. The results in Table 3.3 and Table 3.5 reveal that parental education is responsible for most of the variance and that health contributes the least to the predicted HC estimates. The eigenvalue for PC1 is the only component with an eigenvalue over 1 (estimated as 2.41582). It is retained as the component used to predict HC_p. Furthermore, the proportion of variance explained by PC1 is double that of PC2.

Table 3.5: Factor weights for observed indicators.

Variables	Principal component 1 (PC1)	Principal component (PC2)
Education	0.4449	0.0738
Mothers education	0.5448	-0.0895
Fathers education	0.5458	-0.1264
School quintile	0.4247	-0.1749
Health	0.1642	0.9695
Proportion of variance explained	0.4832	0.1941
Eigenvalue	2.41582	.970373

Source: NIDS wave 5; authors' calculations. Post-stratified weights applied

Thus, considering all the goodness of fit tests and sensitivity analysis estimates, the conclusion is that the model has a good fit and the method used is robust. HC_L per individual is therefore estimated for analysis by gender, race, geographical area and income in South Africa.

3.5.3 Human Capital frequency distribution by subgroup in South Africa

The predicted HC_L variable has an estimated standard deviation of 0.59, a minimum value is -0.68 and maximum value is 1.48. The estimated concentration of individual HC_L is presented in Figure 3.2 using k-density curves. Numerous individuals are estimated as having low levels of HC_L. The sensitivity analysis illustrates the robustness of the results of the concentration of HC_L, given that lower levels of HC continue to dominate the k-density estimations of HC, when applying the adjusted CFA measurement models and the PCA estimation approach. This is expected given the latent measure of correlated indicators maximises common variance and thus the tendency for greater variance in the latent measure than that of a single proxy indicator. However, Wendelspiess Chávez Juárez (2015) identified that when measuring the inequality of a latent construct, factor analysis is more accurate using multiple indicators in comparison to a single proxy, which is subject to proxy measurement error and is downward biased. Parental education is estimated as having a greater impact on the variance of the latent HC in comparison to education attainment and has greater inequality than education attainment (see

Table 3.3); thus the variance of HC_L will be less aligned with education attainment and more aligned with parental education. Regardless of the measurement method applied, few individuals are indicated as having high levels of HC.

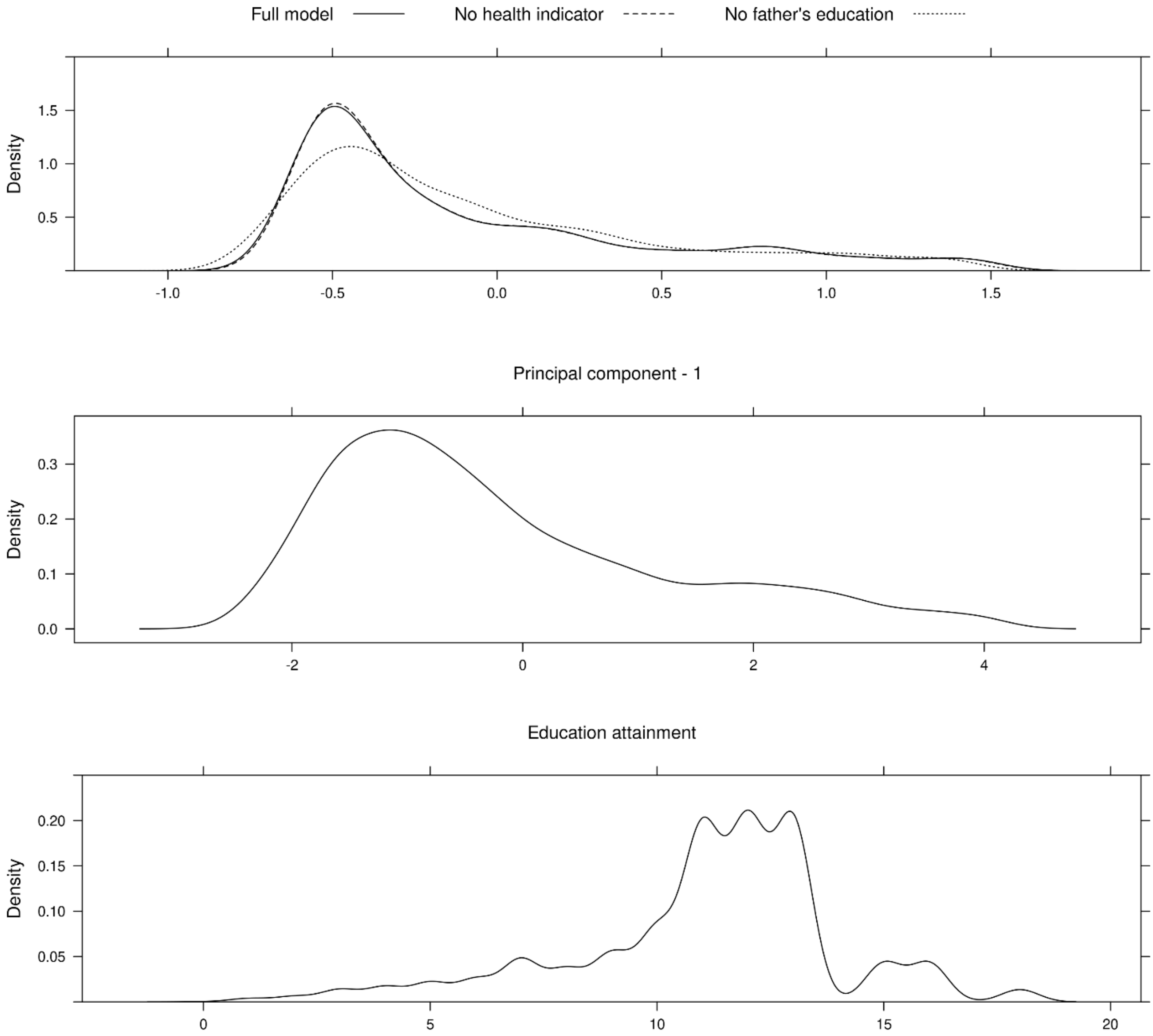


Figure 3.2: K-density curves: HC distribution results.

Source: NIDS wave 5; authors' calculations. Post-stratified weights applied.

The histogram presented in Figure 3.3 below furthermore illustrates a distinct difference in the distribution of HC using the two measurement approaches. A large number of individuals are estimated as having low levels of HC_L, however moderate levels of HC_E.

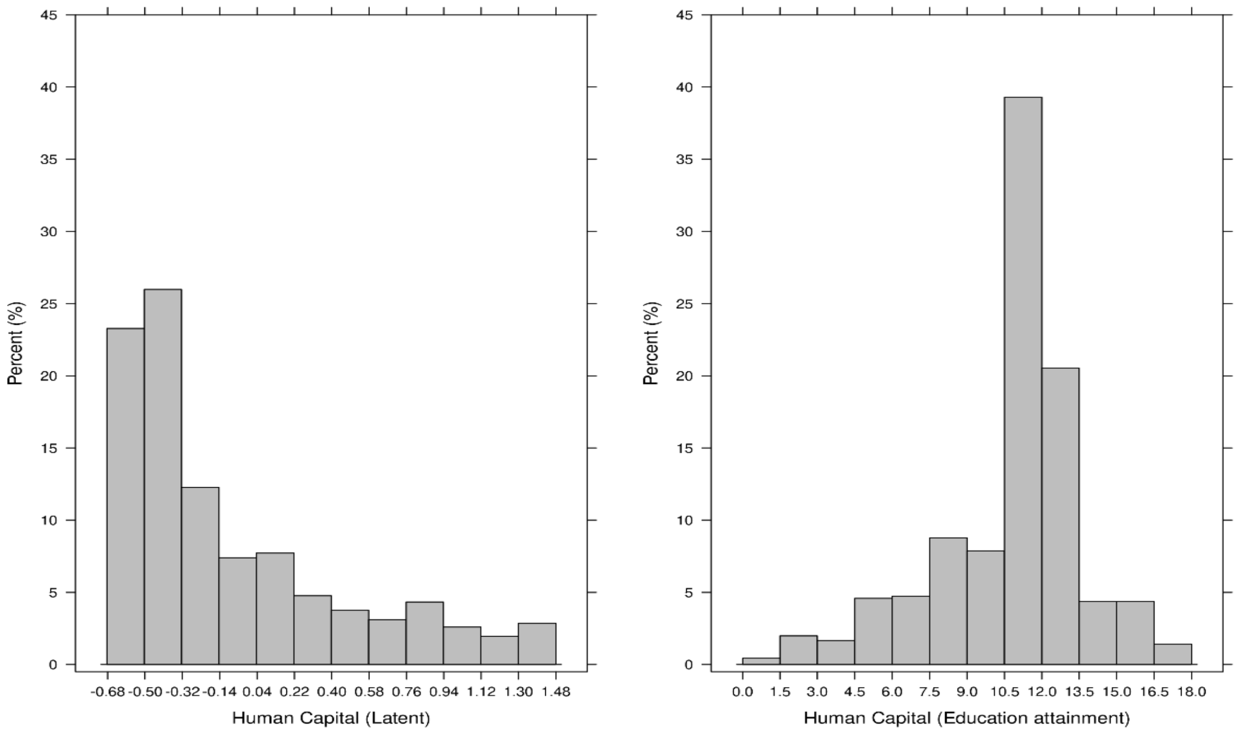


Figure 3.3: Distribution of HC_L and HC_E.

Source: NIDS wave 5; authors' calculations. Post-stratified weights applied.

The distribution profile of HC_L and of HC_E by race, gender, geographical area and income quintile are presented in Table 3.6 and Table 3.7 respectively. In Table 3.6 the HC_L quintiles are calculated by dividing the population into five groups of equal number according to the value of their HC_L. The findings illustrate that the individuals in the sample with the top 20% of HC_L are most commonly in income quintile 5, White, male and resident in urban areas. The subgroup profile of HC_E is presented in Table 3.7 considering five education categories[57]. Individuals in the top HC_E category are predominantly income quintile 5 earners, White, female and resident in urban areas.

[57] HC_L and HC_E estimates are not directly comparable by subgroup using quintile distributions given data limitations. Thus, the subgroup distribution of HC_L is presented by quintile and of HC_E using five education categories, which is the standard approach to estimating rates of return to education in South Africa (Van Broekhuizen, 2011 and Branson and Leibbrandt, 2013)

Table 3.6: HC_L quintiles by gender, race, geographical area and income quintile.

	Human capital (Latent) quintiles					
	1	2	3	4	5	Total
Geographical area						
Urban	15.0 (0.013)	18.5 (0.016)	19.3 (0.015)	22.2 (0.015)	24.9 (0.018)	100
Rural	35.3 (0.023)	24.5 (0.020)	21.3 (0.023)	13.5 (0.016)	5.5 (0.013)	100
Gender						
Female	20.1 (0.014)	19.8 (0.017)	21.1 (0.018)	20.5 (0.016)	18.5 (0.017)	100
Male	20.3 (0.017)	20.3 (0.018)	18.7 (0.018)	19.5 (0.017)	21.1 (0.021)	100
Race						
African	25.1 (0.014)	23.3 (0.015)	23.3 (0.015)	17.2 (0.013)	11.1 (0.014)	100
Coloured	5.7 (0.013)	19.9 (0.040)	17.1 (0.035)	39.2 (0.053)	18.1 (0.044)	100
Asian & Indian	0.0 (0)	16.1 (0.064)	9.5 (0.042)	52.5 (0.110)	21.9 (0.113)	100
White	0.0 (0)	0.0 (0)	0.8 (0.006)	22.9 (0.031)	76.4 (0.031)	100
Income quintiles*						
1	41.4 (0.032)	25.4 (0.033)	12.8 (0.018)	10.6 (0.020)	9.8 (0.033)	100
2	31.5 (0.034)	25.2 (0.031)	21.9 (0.032)	16.9 (0.025)	4.5 (0.019)	100
3	18.7 (0.022)	27.5 (0.028)	25.7 (0.032)	18.5 (0.022)	9.7 (0.019)	100
4	8.5 (0.018)	14.7 (0.026)	23.1 (0.029)	26.8 (0.031)	27.0 (0.035)	100
5	1.2 (0.005)	7.5 (0.017)	15.5 (0.027)	27.1 (0.031)	48.7 (0.035)	100

Notes: Standard errors presented in brackets below percentage estimates. HC_L and income quintiles are calculated by dividing the population into 5 equal groups according to the value of HC_L and income, respectively.

Source: NIDS wave 5; authors' calculations. Post-stratified weights applied.

Considering the distribution of HC by geographical area, it is observed that only 5.5% of rural compared to 24.9% urban employed residents are in the 20% of the sample with the highest HC_L. In contrast, 35.3% of rural dwellers are in the lowest HC_L quintile (compared with only 15%). The HC_E categories confirm the dominance of higher levels of HC in urban areas, however, smaller disparities between urban and rural areas at the lowest and highest levels of HC are evident in Table 3.7.

Table 3.7: HC_E attainment categories by gender, race, geographical area and income quintile.

	Human capital (education attainment) categories					
	Primary or less	Secondary	Matric	Some tertiary	Bachelor degree	Total
Geographical area						
Urban	7.3 (0.009)	31.2 (0.018)	24.0 (0.017)	25.0 (0.017)	12.6 (0.011)	100
Rural	17.7 (0.018)	42.8 (0.024)	16.2 (0.021)	16.0 (0.018)	7.2 (0.015)	100
Gender						
Female	8.8 (0.009)	32.3 (0.019)	19.9 (0.017)	25.8 (0.019)	13.3 (0.014)	100
Male	10.9 (0.013)	35.8 (0.022)	23.9 (0.020)	20.1 (0.019)	9.4 (0.011)	100
Race						
African	11.7 (0.010)	38.4 (0.018)	20.3 (0.015)	22.2 (0.016)	7.4 (0.008)	100
Coloured	9.0 (0.019)	37.2 (0.050)	15.5 (0.037)	23.4 (0.048)	14.9 (0.040)	100
Asian & Indian	3.9 (0.036)	21.7 (0.072)	33.9 (0.110)	16.9 (0.058)	23.7 (0.112)	100
White	0.0 (0)	7.3 (0.021)	34.8 (0.040)	26.3 (0.032)	31.6 (0.034)	100
Income quintiles*						
1	21.7 (0.024)	52.9 (0.035)	18.0 (0.030)	6.8 (0.014)	0.6 (0.002)	100
2	18.0 (0.028)	42.8 (0.035)	22.8 (0.030)	14.3 (0.032)	2.0 (0.011)	100
3	7.9 (0.015)	47.6 (0.032)	24.3 (0.027)	17.5 (0.023)	2.6 (0.008)	100
4	1.9 (0.007)	19.7 (0.028)	28.2 (0.034)	40.1 (0.036)	10.2 (0.018)	100
5	0.4 (0.004)	7.7 (0.017)	16.6 (0.030)	34.9 (0.035)	40.5 (0.033)	100

Notes: Standard errors presented in brackets below percentage estimates.
Source: NIDS wave 5; authors' calculations. Post-stratified weights applied.

Considering the distribution of HC by gender, males are slightly more likely than females to have quintile 5 levels of HC_L. This is in contrast to HC by education attainment, where females are more dominant in the two highest HC_E categories. This is evident with 13.3% of females having obtained Bachelor's degrees or more, in comparisons to 9.4% of males. The differences between the HC estimates by gender does not adequately explain the large differences in labour income earnings between genders. While controlling for several factors, the OLS results presented in Table 3.2 above illustrate that males' incomes are 49% higher than females. Possible reasons for the higher differences in earnings between genders includes women shouldering a disproportionate burden of care work in the HH (Casale and Posel, 2005 and Magadla *et al.*, 2019), gender stereotyping, and a wage ceiling, given that women are restricted from participating in the highest paid positions in the economy (Mosomi, 2019).

The racial HC_L distribution differences are extreme; 0% of Whites but 50% of Africans fall in the two lowest HC_L quintiles. The estimates confirm how the unequal development of HC under the apartheid-era continues to disadvantage other race groups (particularly Africans) in comparison to White. The disparities between race groups are however not as distinct when considering HC_E, with 57.9% of Whites in the top two HC_E categories in comparison to 76.4% in the top HC_L quintile alone. The comparable measure for Africans are 29.6% for the top two categories of HC_E and only 11.1% in the top category of HC_E.

Both measures clearly demonstrate the importance of human capital in explaining income distribution in SA. 66.8% of income quintile 1 and 56.7% of income quintile 2 have the 2 lowest quintiles of HC_L. The corresponding shares for HC_E are 74.6% and 60.8%. In contrast 75.8% and 75.4% of income quintile 1 have the highest quintile of HC_L and HC_E respectively. Human capital therefore clearly matters. This is expected, given other findings of higher earnings returns for higher levels of education in South Africa (Branson and Leibbrandt, 2012 and van Broekhuizen, 2011).

The CFA measurement model specified thus identifies the continued disadvantages between racial groups, given that parental educational attainment and school quality are incorporated as indicators of HC_L. This provides evidence that Whites will continue to remain advantaged and Africans disadvantaged unless the latter are afforded the opportunity to acquire better quality education and improved education attainment at the higher levels. Racial differences in schooling quality remain evident in South Africa and without high quality education for all citizens, the cycle of poverty and economic inequalities will persist.

Considering the profile of HC by income quintile identifies the benefit of profiling individuals by population quintile, which HC_E does not accurately allow. For example, it is expected that the concentration of income quintile 1 individuals will steadily decrease from the lowest to the highest HC_E category as per the HC_L quintile profiling. However, this is not the case; there is an increase from HC_E category 1 to category 2. Furthermore, for income quintile 4 it is expected that there should be a steady increase in the percentage of employed individuals from education category 1 to 5 (as per HC_L quintile). However, there is a reduction in the percentage of individuals in education category 5 compared to education category 4. The reason for these unexpected differences is expected to be a result of the number of individuals in the different education categories. The education category 2 has the highest percentage of individuals in the sample at 36.2% in comparison to education category 5 which has the lowest percentage at 10.14%. To investigate the relationship between HC_L and earnings, further research on the earnings returns to HC_L is recommended.

3.6 Conclusion

HC is a complex concept to measure, given that it is unobserved and that there is no clear definition or specified indicators, to capture its meaning. Education has thus become generally accepted as a proxy for HC, mainly because there is extensive data available. However, a latent-variable approach has been proposed, since HC is in fact a multidimensional unobserved construct and the potential for proxy measurement errors. The authors employed a reflective CFA measurement model to estimate the latent construct HC and the model was identified as having a good fit and robust.

The observed indicators included in the model were selected based on empirical evidence of high returns to education attainment, inequalities in education quality, intergenerational transmission of poverty and inequality in South Africa and the contribution of health as a factor impacting education attainment and labour market outcomes. The indicators were furthermore in accordance with social reproduction theory and microeconomic HC theory, and sensitivity analyses were conducted to analyse the robustness of the results. HC profiles by HC_L quintiles and HC_E categories are presented.

Valuable insights are gained from estimating HC_L in a CFA measurement model and presenting the HC_L and HC_E profiles by SES indicators. Firstly, it was found that parental education is associated with the largest amount of variance in HC, while the health indicator captures the least. HC_L is noted as illustrating a greater frequency distribution in lower levels of HC in comparison to HC_E. The differences are highlighted because of the inclusion of parental education and school quality in HC_L, which captures the SES of HHs. Furthermore, the HC_L variable allows for analysis by population quintiles, which the HC_E variable does not.

Thus, it is suggested that without improvements in schooling quality, valuable education attainment and parental education for those disadvantaged during the apartheid-era and subject to poor service delivery post-apartheid, will remain low and thus inequalities in HC and income will persist. In addition, the results provide evidence that the social reproduction theory is evident with the education system in South Africa perpetuating social stratification. The sensitivity analysis confirmed the robustness of the results, with model fit improved when removing health as an observed indicator.

Despite the reflective model providing insights into the analysis of HC in South Africa, the process of measuring HC using a latent-variable approach in this study highlights why the education-based approach to measuring HC persists. The reasons include data availability, simplicity of the approach and that the HC_E mostly presents expected distributional differences between subgroups. The latent-variable approach however allows for population quintile analysis and multiple indicators to be incorporated so as to account for the South African context of the intergenerational transfer of poverty and inequality and a schooling system which perpetuates this cycle. For future research, a SEM framework is recommended to further analyse the insights gained in estimating HC_L, and to develop a model which allows for the estimation of the causal paths between HC and the labour market. Furthermore, research focused on understanding health as an indicator of HC in South Africa is also recommended.

3.7 Appendices

Appendix 3.A: Correlation matrix of study variables.

	Education	Mother's education	Father's education	School quintile	Health	Labour income	Race	Geographical area	Gender
Education	1.000								
Mother's education	0.4419	1.000							
Father's education	0.4478	0.6858	1.000						
School quintile	0.2834	0.4200	0.4280	1.000					
Health	0.1480	0.1343	0.1103	0.0783	1.000				
Labour income	0.4511	0.3338	0.3458	0.2929	0.0644	1.000			
Race	0.3000	0.4600	0.4786	0.6110	0.0446	0.3843	1.000		
Geographical area	0.1923	0.2112	0.2546	0.2410	0.0085	0.1328	0.2443	1.000	
Gender	-0.0892	0.0413	0.0240	-0.0326	0.0301	0.1386	0.0199	0.0281	1.000

Source: NIDS wave 5; authors' calculations. Post-stratified weights applied.

Appendix 3.B: CFA measurement model: Standardized coefficient estimates, goodness of fit results and predicted HC estimates (unweighted).

	Effect	Estimate	S.E.	P-value
HC ~	Education	0.549	0.015	0.000
	Mother's education	0.825	0.011	0.000
	Fathers education	0.830	0.011	0.000
	Health	0.134	0.020	0.000
	Quintile	0.490	0.016	0.000
Goodness of fit results*	Test	Benchmark	Result	Good fit
	CD	Close to 1	0.836	Yes
	CFI	≥ .90	0.994	Yes
	TFI	≥ .90	0.987	Yes
	SMRM	≤ .06	0.019	Yes
	RMSEA	≤ .08	0.038	Yes
	Chi-squared	p-value <0.05	0.00*	No

Notes: *Rejecting the null hypothesis indicates poor model fit.

Source: NIDS wave 5; authors' calculations.

CHAPTER 4

4 RECONCILING THE MEANING AND MEASUREMENT OF HUMAN CAPITAL AS A DETERMINANT OF SOCIOECONOMIC INEQUALITIES[58]

Abstract

The Human Capital Index (HCI) developed by the World Bank (2018a) provides a measure which can be used to study human capital (HC) productivity gaps between countries. The HCI uses measures of survival, education and health to estimate, at a country level, the HC "a child born today can expect to attain by her/his 18th birthday, given the risks of poor health and poor education where she lives" (World Bank, 2018a:3). The socioeconomic disaggregated human capital index (SES-HCI), an extension of the HCI, provides a means for analysing HC inequalities within countries. This study estimates SES-HCIs for South Africa by income quintiles, school quintiles, geographical area, gender and race. The main driver of HC inequalities in all the SES indicators is found to be the quality of schooling. Factors to address the inequalities and the limitations of the measuring instruments are identified.

Keywords: human capital; inequality; human capital index

[58] The article has been published online through the international journal Development Southern Africa. The co-authors are Prof Keeton and Prof Rogan. Available at: https://doi.org/10.1080/0376835X.2021.1941779

4.1 Introduction

The Human Capital Project (HCP), of the World Bank (2018a) introduced a Human Capital Index (HCI) for all countries, in order to encourage better investments into human capital (HC) by identifying national productivity gaps. For the purpose of the HCI, the World Bank accepted the broad definition of HC, stated by the OECD (2001:18) as "the knowledge, skills, competencies and attributes embodied in individuals that facilitate the creation of personal, social and economic well-being". The HCI did not measure all aspects contained in this definition. Instead, it used just three components, survival, education and health measured at a country level, to estimate the HC "a child born today can expect to attain by her 18th birthday, given the risks[59] of poor health and poor education where she lives" (World Bank, 2018a:3). Proxies used to measure these three components were: for survival, the under-5 mortality rate (U5MR); for education, expected learning-adjusted years of schooling; and for health, adult survival and stunting rates.

The HCI results indicated that South Africans born today can, on average, expect to obtain 41% of their full potential health and education, by the age of 18 years. South Africa ranked 126th out of 157 countries measured (World Bank, 2018a), which confirms the need for further research for understanding the multidimensional nature of the low levels of HC in South Africa. South Africa is furthermore one of the most socioeconomically unequal countries in the world with a gini coefficient for income[60] of 0.63 in 2014 (World Bank, 2019a) and 0.95[61] for wealth[62] (Orthofer, 2017). Leibbrandt *et al.* (2012) demonstrated that income inequality in South Africa is driven by the labour market, and so to reduce income

[59] It is the risks faced based on an assumption that "they experience currently-prevailing risks of poor health and poor education faced by children aged 0-17" (D'Souza *et al.*, 2019)

[60] The National Income Dynamics Study (NIDS) calculates household income as the sum of labour market income (net of taxes), government grants, other government income, investment income, remittances received, subsistence agriculture income and imputed rent for owner-occupied housing (Chinhema *et al.*, 2016). For the purpose of calculating inequality, the household income is divided by the number of household residents to obtain a household per capita income, which is assigned to each household member (Leibbrandt *et al.*, 2012).

[61] HC was not included in the measure of wealth (Orthofer, 2017)

[62] Wealth (also known as 'net worth') is calculated as the difference between the market value of all assets and liabilities (Orthofer, 2017) and in NIDS household wealth is calculated by summing net financial wealth, net business equity, net real estate equity, value of vehicles, total value of pension/retirement annuities and livestock wealth (Chinhema *et al.*, 2016).

inequality policies to effectively grow skills and productivity are required. The bidirectional relationship between HC and income inequality (Chani e*t al.*, 2014), has thus resulted in policy commitments, as stated in the National Development Plan (NDP) 2030, to develop HC through increasing investments and furthermore move to a knowledge-intensive economy (NPC, 2012).

However, because of data and methodological challenges, analyses to date on the impacts on inequality of HC, measured using a multidimensional approach, have been limited. Most researchers in South Africa have used educational attainment to measure HC (Lam, 1999; Van der Berg, 2010; Pellicer and Ranchhod, 2012; Branson and Leibbrandt, 2013 and Wanka, 2014). For a discussion on the problems of this narrow approach and alternative methods of measuring HC see Oxley *et al.* (2008) and Folloni and Vittadini (2010).

The World Bank's (2018a) HCI does not provide any indication of the distribution of HC, or information needed to analyse the causes of national HC and productivity gaps and track progress made through policy intervention. The World Bank has begun to address this limitation by extending the HCI methodology to construct indices of HC disaggregated by socioeconomic status (SES) quintiles within countries (D'Souza *et al.*, 2019). The SES-disaggregated human capital index (SES-HCI) has been estimated for 51 countries (of which South Africa is not one), approximately a third of the countries in the HCI study. The limitations of available data are cited as the main reason for the exclusion of the remaining countries.

Liu and Fraumeni (2021) discusses three complementary HC indexes developed namely the United Nation's Human Development Index (HDI)[63], the Institute for Health Metrics and Evaluation Human Capital Index (IHME HCI)[64] and the World

[63] The HDI was developed to measure human development as a composite index and included in the first Human Development Report written by a team at the UNDP led by Mahbub ul Haq (UNDP, 1990). South Africa ranks 114th out of 189 countries (UNDP, 2020)
[64] The IHME HCI measures the "expected years lived from age 20 to 64 years and adjusted for educational attainment, learning, and functional health status" (Lim *et al.*, 2018: 1217). South Africa ranks 144th out of 195 countries.

Economic Forum's Global Human Capital Index[65] (WEF GHCI). The indexes use various macroeconomic education and health indicators as the two basic dimensions of HC towards estimating a counties' distance from optimal HC (the distance from maximum HC attainable).

In the South African context, Fransman and Yu (2018) derived a Multidimensional Poverty Index (MPI) disaggregated by gender, race and spatial location, revealing that the MPI of Africans contributed more than 95% to multidimensional poverty in South Africa. The three main indicators of the MPI driving poverty in South Africa were identified as the number of years of schooling, unemployment and disability. Within this context, this study aims to measure the education and health components impacting the development of a child's potential and resultant productivity gaps in South Africa by SES indicators using the SES-HCI.

This allows recommendations to be made on how the SES-HCI can best be used to measure and monitor progress made by investments in HC, to ensure that all children reach their potential, in the South African context of very high inequality.

4.2 Materials and Methods

4. 2.1 Data and Sample

Primary and secondary data used for this study are from Wave 5 of the 2017 National Income Dynamics Study (NIDS) dataset and the South African Demographic Household Survey (SADHS) 2016. The NIDS is a panel dataset compiled biennially by the South African Labour and Development Research Unit (SALDRU) based at the University of Cape Town (UCT). The first wave was released in 2008, administered to a nationally representative sample, and the most recent wave, Wave 5, was conducted in 2017 (see Brophy *et al.*, 2018). It comprises a total

[65] The WEF GHCI has been constructed since 2015 with the updated methodologies and most recent estimates presented for 130 economics for 2017 in The Global Human Capital Report 2017 (World Economic Forum, 2017). South Africa ranks 87th out of 130 countries.

of approximately 37000[66] individuals in 13719 households as a cross-sectional dataset. Post-stratification weights as per SALDRU (2017) were applied.

Two main reasons for using the NIDS were the very comprehensive nature of the data and the fact that this research forms part of a larger study which analyses HC using different measurements and the impact of HC on inequality. Using the same dataset will make future alternative HC measurements comparable. However, the Wave 5 NIDS data has insufficient observations for estimating the U5MR, which the HCI requires to calculate survival (only 26 observations of under-5 child deaths were recorded). Therefore, child mortality rates from a document by the NDoH, StatsSA, SAMRC and ICF (hereafter referred to as NDoH *et al.*, 2019) on U5MRs from the 2016 SADHS[67] dataset were used as part of the component of survival to supplement the NIDS data. The fact that the time periods of the two studies differ is acknowledged, but is unlikely to impact the results. The NDoH *et al.* (2019) results were concerned with the previous 5 or 10 years, whereas the NIDS measured observations over the previous 12 months. Also, the SADHS reference year was 2016 while that for NIDS was 2017.

4.2.2 Index framework and variables

The HCI and SES-HCI methodologies were published in Kraay (2018) and D'Souza *et al.* (2019) respectively. The SES-HCI followed largely the same method as the HCI, but at a subgroup level based on a variety of SES indicators, namely: income, race, gender, geographical area and school quality. Variables were included based on data availability as suggested in D'Souza *et al.* (2019).

[66] The total numbers of individuals successfully interviewed in Wave 1 and Wave 5 were 26776 and 37 368 respectively (Brophy *et al.*, 2018).

[67] The Demographic Household Survey (DHS) program has fielded over 400 surveys across 90 countries, collecting household demographic data including birth histories. In South Africa, three surveys have been conducted: in 1998, 2003 and 2016. The SADHS reviewed 11 083 households (NDoH *et al.*, 2019).

The HCI index is formulated as:

$$HCI = Survival\ x\ School\ x\ Health \tag{1}$$

With the three components defined as:

$$Survival = \frac{1 - under\ 5\ mortaltiy\ rate}{1} \tag{2}$$

$$School = e^{\emptyset\left(expected\ years\ of\ school\ x\ \frac{Harmonized\ Test\ Score}{625}\ -14\right)} \tag{3.1}$$

$$Health = e^{(\gamma_{ASR}\ x\ (Adult\ survival\ rate-1)+(\gamma_{Stunting}\ x\ (Not\ stunted\ rate\ -1)/2)} \tag{4.1}$$

where under-5 mortality rate (U5MR) is defined as the number of deaths under age 5 per 1000 births in the last year. $\emptyset$ is the rate of return to schooling which in the World Bank study (Kraay, 2018) is 8%[68]. In equation 3.1 the expected years of schooling (EYS) are adjusted for a harmonised test score (HTS) calculated from the international TIMSS[69] and PIRLS[70] assessments (maximum score of 625 used to account for schooling quality), to calculate a learning-adjusted years of education. γ_{ASR} is the return to an additional unit of adult survival and $\gamma_{stunting}$ the return to not being stunted, which Kraay (2018)[71], for the purpose of estimating the HCI, estimated as 0.65 and 0.35 respectively. The interpretation of the health return estimates is that when the health component improves to the extent that the adult survival rate (ASR) increases by one percentage point, then worker productivity increases by 0.65 percentage points (Kraay, 2018).

[68] In the South African context, a standard 8% earnings return to education may underestimate the returns, with Depken *et al.* (2019) highlighting approximately 18% as an accurate estimate of returns per year of education attained in South Africa. The estimation approach included by Depken *et al.* (2019) included ordinary least squares (OLS) and instrumental variables (IV). Furthermore, in comparison to primary school attainment, van Broekhuizen (2011) estimated high and increasing rates of return to higher education, noting that secondary schooling, matric, Bachelor's degree and post-graduate degree attainment had 13%, 35%, 115% and 180% average earnings returns relative to primary schooling respectively.

[69] Trends in International Mathematics and Science Study (TIMSS) is an IEA assessment which was administered to Grade 9s in South Africa in 1995, 1999, 2003, 2011 and 2015 (van der Berg and Gustafsson, 2019). To calculate harmonised test scores, TIMSS 2015 results available from Reddy *et al.*, 2016, were used.

[70] Progress in International Reading Literacy Study (PIRLS) is an IEA assessment which was administered to Grade 4s in South Africa in 2006, 2011 and 2016 (van der Berg and Gustafsson, 2019). To calculate the harmonised test scores, PIRLS Literacy 2016 results available from Howie *et al.* (2017) were used.

[71] See Kraay (2018:41-42) for details on the literature used to calculate the returns used for health, and Kraay (2018:34-36) for education.

Similarly, an improvement in overall health associated with a reduction in stunting rates of 10 percentage points raises worker productivity by 3.5 percent (D'Souza *et al.*, 2019). Where data for both stunting rates and ASRs are available, the average of the improvements in productivity associated with both health measures is applied in the health component of the HCI, as noted in equation 4.1. In equation 3.1, 14 is the maximum number of years of school enrolled for from age 4 to 17. The reason for including the returns to education and health in the HCI calculation is that it allows one to estimate how much a reduction in full education or health reduces earnings and productivity (Kraay, 2018). The values selected for the returns to education[72] and health by Kraay (2018) are anchored in an extensive study of the empirical literature.

This study uses methodological elements from both the SES-HCI and HCI to estimate the best context-specific SES-HCIs for South Africa. EYS, in equation 3.1, was also estimated using both educational attainment and enrolment (Woßmann, 2003). Education enrolment has become near universal in South Africa (StatsSA, 2011), thus both attainment and enrolment were used to provide comparable measures of EYS. Results are presented in Table 4.3 below. For education attainment, the definition for EYS is the mean years of schooling an individual could expect to obtain by and including the age of 18 years[73].

Because income inequality in South Africa is largely driven by the labour market (Leibbrandt *et al.*, 2012) and education attainment drives labour market outcomes (Taylor, 2019 and Kamanzi *et al.,* 2021), when measuring the SES-HCI disaggregated by income, the school component was adjusted to account also for tertiary education. The age 25[74] is used as the expected age at which education attainment will have been completed (see UNDP, 2020). To calculate mean years of education attainment by age 25, the NIDS highest education attainment variable was recoded as per Branson and Leibbrandt (2013) and 16 years the potential

[72] The 8% rate of return to schooling suggested by Kraay (2018) is deliberately chosen to be on the lower range, given a vast majority of the returns to education estimated do not control for health.
[73] Age 18 is the expected age at which a child will complete grade 12 (Matric) in South Africa. Enrolment is measured until age 17 for the HCI but, given that enrolment at age 17 would result in attainment at age 18, age 18 is included when using the attainment approach.
[74] The HCI definition would need to be adjusted from expected HC at age 18 to age 25.

maximum attainment. For secondary schooling values, the years of schooling were adjusted by the relevant HTS per school quintile on a graded curve, to obtain learning-adjusted years of education, before applying the return to education. Years of tertiary education were not adjusted by the HTS. The school component for the SES-HCI adjusted to include tertiary education is presented as equation 3.2 and the results are shown in Table 4.8.

$$School = e^{\emptyset\left(\left(expected\ years\ of\ school\ x\ \frac{Harmonized\ Test\ Score}{425}\right) + mean\ years\ of\ tertiary\ education\ -16\right)} \qquad (3.2)$$

To account for education quality in the disaggregated SES indicators, average HTS per school quintile[75] (using PIRLS (2016), TIMSS Mathematics (2015) and TIMSS Science (2015)) from Reddy *et al.* (2015) and Howie *et al.* (2017) are used and shown in Table 4.1. Quintile 5 (Q5) schools have a mean HTS of 425 (68% of 625) and quintile 1 (Q1) schools 315 (50% of 625). With an average mean score of 343 (54.9%) and 345 (55.2%) out of 625 calculated using World Bank (2018b) and NIDS data respectively, South African HTSs are low overall in the global context. The country with the highest HTS was Singapore with 581 (World Bank, 2019c). Singapore also had the highest HCI score: 0.88. Patrinos and Angrist (2018) note that the large difference between the scores of Singapore and South Africa is more than 2 standard deviations and is equivalent to a generation of lost schooling.

To investigate the distributions differences in HC in South Africa, the HTS is also graded on a curve to illustrate the difference within the country as per the SES-HCI. Results using a maximum HTS of 425 will therefore also be presented, thus applying the quintile 5 schools mean HTS as the maximum score. The calculated

[75] School quintile (1 to 5) is the school quality variable used in the NIDS data. For each income quintile, gender, race and geographical area a harmonised test score was calculated using a score per school quintile calculated by averaging the values for the TIMSS Maths and Science scores and PIRLS scores to provide a score per school quintile as found in Reddy *et al.* (2016) and Howie *et al.* (2017). Thereafter the percentage of the school going population per disaggregated variable attending that particularly school quintile (available from the NIDS data) was calculated and a test score estimated. See the results in Table 4.1. The national figure obtained calculated from this method was 345 and the figure provided by the World Bank (2018b) for South Africa was 343. Thus, the scores calculated are expected to provide a reliable estimate given the available data.

HTSs for the different SES indicators are presented below when discussing the results for the SES-HCI by geographical area, gender and race.

Table 4.1: South African harmonised test score per school quintile

School Quintile	PIRLS (2016)	TIMSS MAT (2015)	TIMSSSCIE2015	Avg/avg student*
1	288	341	317	315.33
2	299	341	317	319
3	303	341	371	338.33
4	328	423	425	392
5	426	423	425	424.67

Notes: * This is the average score across the PIRLS (2016), TIMSS MAT (2015) and TIMSSSCIE 2015 per student and school quintile.

Source: Reddy *et al.* (2015) and Howie *et al.* (2017)

To measure stunting, the NIDS provides height for age *z*-scores for children under-5. Stunting is measured as being 2 standard deviations below the median height for age scores, as used by studies on stunting based on NIDS data (see Otterbach and Rogan, 2017). To estimate the health component, the probability of not being stunted per subgroup was adjusted by the return to reduced stunting provided in Kraay (2018).

Following D'Souza *et al.* (2019), this study omits adult survival from the health component when calculating the SES-HCI[76]. Thus, the health component presented in equation 4.1 is adjusted to:

$$Health = e^{\gamma_{Stunting}\, x\, (Not\ stunted\ rate-1\,)} \tag{4.2}$$

SES indicators used to disaggregate HCI in this study are income quintile[77] (the NIDS derived household income variable was divided by household size to calculate HH income per capita), school quintile (Q1 – Q5, as categorised in the NIDS dataset), geographical region (urban versus rural[78]), gender and race. Each of these elements is related to earnings inequalities in South Africa (StatsSA, 2019). Details of the

[76] Kraay (2018) defined ASR as the fraction of 15-year olds surviving until age 60.
[77] The NIDS dataset provides a derived household income variable which was divided by household size to calculate a household income per capita which was used to create 5 data samples, one for each quintile.
[78] Traditional and farm options were combined to generate a rural variable.

different proxies and data applied for each of the components of the SES-HCI indicators are presented in Table 4.2. Note that when disaggregating by school quintile it is not possible to calculate an under-5 stunting or mortality rate because most children under 5 years of age are not in school yet and therefore cannot be allocated to a school quintile. Instead, and because higher school quintiles require higher fees, the relevant income quintile values were used as indicators for the U5MRs and health components. Changes in under-5 mortality and stunting rates are expected to move in the same direction.

Lastly, D'Souza *et al.* (2019) noted that, since difference datasets were employed, interpreting and comparing HCI results using different data sources, as in Kraay (2018) and D'Souza *et al.* (2019), should be undertaken with caution.

Table 4.2: Study variations by data and survival, education and health components.

Component	Data	Survival*	Education				Health	
			Age category	*Expected years of schooling (EYS)*	*Harmonised test score (HTS)*	*Rates or return*	*Stunting*	*ASR****
HCI (Kraay, 2018). World Bank, 2019c)	National Administrative data[79]	Under-5 mortality rate (U5MR)	4 – 17	Enrolment rates	TIMSS and PIRLS	8%	Under-5 stunting rate (U5SR)	ASR
SES-HCI (D'Souza *et al.,* 2019)	Household data (DHS; MIC)	U5MR	6-17	Enrolment rates	TIMSS and PIRLS	8%	U5SR	___
SA SES-HCI by income. Table 4.4, 4.5 & 4.7	Household data (NIDS; DHS).	U5MR	4-17	Enrolment rates	Table 4.6	8%	U5SR	___
SA SES-HCI by income (with tertiary education). Table 4.8.	Household data (NIDS; DHS)	U5MR	4-25	Education attainment	Table 4.6	8%	U5SR	___
SA SES-HCI by race. Table 4.10, 4.11 & 4.12.	Household data (NIDS; DHS)	U5MR	4-17	Enrolment rates	Table 4.9	8%	U5SR	___
SA SES-HCI by region. Table 4.10, 4.11 & 4.12	Household data (NIDS; DHS)	U5MR	4-17	Enrolment rates	Table 4.9	8%	U5SR	___
SA SES-HCI by gender. Table 4.11, 4.12 & 4.13	Household data (NIDS; DHS)	U5MR	4-17	Enrolment rates	Table 4.9	8%	U5SR	___
SA SES-HCI by school quintile.** Table 4.13	Household data (NIDS; DHS)	U5MR proxy**	4-17	Enrolment rates	Table 4.1		U5SR	___

Notes: * For the SES-HCI estimates, the total number of observed under-5 deaths using the NIDS was only 26. Thus SES-HCI disaggregated indicators were in addition calculated using the SADHS U5MRs (NDoH *et al.*, 2019). **With regards to disaggregation by school quintile the U5MR per income quintile were used as a proxy in the case of the NIDS data and wealth quintile (calculated using wealth indices estimated by means of Principal Component Analysis) for the SADHS data. Under-5 stunting rates per income quintile was used as a proxy for under-5 stunting per school quintile. ***Adult survival rate.

4.3 Results

With high levels of income inequality in South Africa, inequality between fee and non- or low-fee paying education and healthcare services (Spaull, 2019 and StatsSA,

[79] See Kraay. (2018) footnote on page 1 for all the contributors to the data collection process.

2019) and the bidirectional relationship between HC and earned income (Chani *et al.*, 2014), it is expected that SES-HCI calculations per socioeconomic indicator should show large variations. According to StatsSA (2019), higher income quintiles, quintile 5 schools and urban regions have higher earnings and are thus expected to have higher SES-HCI measures than low-income quintiles, quintile 1 schools and rural regions, respectively. Furthermore, based on the results from the World Bank's (2018a) measure of HCI for South Africa, women (HCI of 0.41) are expected to have only a slightly higher SES-HCI score than men (HCI of 0.36).

However, StatsSA (2019) reported that men have higher levels of employment and earnings and thus it appears that factors other than expected HC at the age of 18 years affect labour market outcomes with respect to gender. The SES-HCI will thus be disaggregated by gender using the NIDS to further investigate the distribution of HC between men and women by the age of 18. With regards to race, and due to the country's history of racial segregation, it is expected that White South Africans would have the highest SES-HCI in comparison to the other race groups (StatsSA, 2019).

To contextualize the results obtained using the NIDS household data, the national HCI calculation using the NIDS dataset is compared to the World Bank's (2019b) estimates in Table 4.3. The impact of substituting the survival to age 5 estimate with SADHS data is also shown. The impact of using education attainment as opposed to enrolment is also demonstrated.

Table 4.3: HCI, South Africa: NIDS data versus World Bank (WB) data calculation.

Component	WB	NIDS	NIDS/SADHS	NIDS
	*Enrolment**	*Enrolment**	*Enrolment**	*Attainment***
Survival to Age 5 (A)	0.96	0.97	0.96	0.97
Expected years of school	9.4	12.4	12.4	10.1
HTS	343	345	345	345
School (B)	0.49	0.56	0.56	0.60
Not stunted	0.73	0.79	0.79	0.79
Adult Survival Rate (ASR)	0.68	-	-	-
Health (C)	0.86	0.93	0.93	0.93
HCI (AxBxC)	0.41	0.51	0.50	0.54

Notes: *Expected years of schooling estimated using repetition adjusted enrolment rates with a potential maximum of 14 years. **Expected years of schooling estimated using education attainment, with a potential maximum of 12 years.

Source: NDoH *et al.*, (2019), World Bank (2019b) and authors own calculations.

The HCI estimates, as per the NIDS data, are all higher than the World Bank's results. The main contributors to this higher measure are the EYS measures and the calculated probability of not being stunted. These result in a seven percentage point increase in both the school component (when using enrolment, but 11 percentage point when using attainment) and the health component~~s~~.

The difference between the estimates of the probability of not being stunted can simply be attributed to dataset differences. The NIDS under-5 sample is smaller than the World Bank's datasets and thus the health component of the HCI calculated by the World Bank (2018b) is likely to be more accurate. Moreover, the World Bank also includes the ASR which could not be calculated from the NIDS dataset. The large difference between the World Bank (2018b) and the NIDS data estimates for EYS, using repetition adjusted enrolment rates, of 9.4 years and 12.4 years respectively, is more difficult to explain unless it is seen as also being due to dataset differences. Which measure is more accurate is unclear. The EYS value for the attainment variable is 10.1 years and is lower than the NIDS enrolment figure. Attainment has a potential maximum of 12 years, versus enrolment, which has a maximum of 14 years, so the values calculated using the NIDS are similar (88.6% for enrolment and 84.2% for attainment). Because the results are similar, enrolment is used to measure the SES-HCI estimates below, the same method used by the World Bank (2018b). Furthermore, with school attendance almost universal, the differences in HC educational outcomes in South Africa are likely to come from school quality (measured by the HTS) and tertiary education attainment, rather than lower than tertiary education attendance or attainment.

4.3.1 Disaggregated index scores by income

The results from the SES-HCI estimates disaggregated by income quintile in Table 4.4 show quintile 1 with an SES-HCI of 0.48 and quintile 5 with an SES-HCI of 0.60, a 12 percentage point difference. Using the HCI definition, this would thus mean that, by the age of 18, the productivity of children born and raised in income quintile 5 households could expect to be 12 percentage points greater than that of children in income quintile 1 households, as a result of the different health and education services received. Surprisingly, quintile 4 has a lower probability of survival to age 5 than quintile 3. While the number of under age 5 deaths decreases consistently from quintile 1 to 5, so do the number of births. This results in a lower survival rate for quintile 4.

Table 4.4: SES-HCI, South Africa: Income quintiles.

Component	Overall	1	2	3	4	5
Survival to Age 5 (A)	0.97	0.97	0.97	0.98	0.95	1.0000
Expected years of schooling	12.4	12.0	12.0	12.5	12.7	13.2
Harmonised test score	345	331	340	345	361	386
School (B)	0.56	0.54	0.55	0.57	0.59	0.62
Not stunted	0.79	0.72	0.79	0.82	0.86	0.88
Adult stunting rate	-	-	-	-	-	-
Health (C)	0.93	0.91	0.93	0.94	0.95	0.96
HCI (AxBxC)	0.51	0.48	0.50	0.52	0.53	0.60

Source: Authors own calculations.

The results of the SES-HCI recalculated using the survival to age 5 rates provided by the larger SADHS dataset, are shown in Table 4.5 (NDoH *et al.*, 2019). SADHS results confirmed increasing under-5 survival rates per wealth quintile, with the exception of the under-5 survival rate for quintile 4 being higher than quintile 5. The under-5 deaths per 1000 births from poorest to the richest quintile were noted as 67, 52, 51, 37 and 41 (NDoH *et al.*, 2019).

Table 4.5: SES-HCI, South Africa: Income quintiles (using survival to age 5 estimates from the SADHS per wealth quintile as a proxy).

Component	Overall	1	2	3	4	5
Survival to Age 5 (A)	0.96	0.93	0.95	0.95	0.97	0.96
Expected years of schooling	12.4	12.0	12.0	12.5	12.7	13.2
Harmonised test score	345	331	340	345	361	385
School (B)	0.56	0.54	0.55	0.57	0.59	0.62
Not stunted	0.79	0.72	0.79	0.82	0.86	0.88
Adult stunting rate	-	-	-	-	-	-
Health (C)	0.93	0.91	0.93	0.94	0.95	0.96
HCI (AxBxC)	0.50	0.46	0.48	0.51	0.54	0.57

Source: Authors own calculations and NDoH *et al.* (2019).

The impact on the HCI results of using the different survival to age-5 estimates is very small. While the SES-HCI calculations in Table 4.5 are slightly lower for all income quintiles, with the exception of quintile 4, when using the SADHS data (NDoH *et al.*, 2019), the difference between income quintiles 1 and 5 changes by only 1 percentage point. This is because the NIDS survival to age-5 estimates, despite the small sample size, are very close to those calculated from the 2011 Census data (StatsSA, 2011)

The HTSs for income quintiles in Tables 4.4 and 4.5 are calculated using the mean HTS per school quintile in Table 4.1 and the percentage of scholars attending the different school quintiles in each income quintile. This is shown in Table 6.

Table 4.6: Average harmonised test score (HTS) calculated using percentage of learners from each school quintile in each income quintile.

School quintile	HTS	Income quintile					
		1	2	3	4	5	National
Quintile 1	315.33	29.88	23.41	20,28	12.58	9.79	22.08
Quintile 2	319	32.43	27.6	24,95	17.37	6.24	24.69
Quintile 3	338.33	28.03	29.75	29,99	27.45	20.8	29.03
Quintile 4	392	7.09	13.25	16,83	23.24	13.27	13.20
Quintile 5	424.67	2.56	5.99	7,94	19.36	49.91	11.01
TOTAL		100	100	100	100	100	100
Average score		331	340	345	361	385	345

Source: Authors own calculations.

According to the NIDS data, only 49.91% of children in income quintile 5 households attend fee-paying (quintile 5) schools. This surprising result is probably explained by the fact that because of South Africa's high income inequality, income quintile 5 starts with households whose per capita income is as low as R4 600 per month[80]. Fees for quintile 5 schools may therefore be unaffordable to many households classified as income quintile 5. This mix of school attendance reduces the differences in HTSs per income quintile. To show greater HC distributional differences of school quality, the HTSs are graded on a curve, with 425 set as the maximum grade in Table 4.7.

Table 4.7: SES-HCI, South Africa: Income quintiles, using a graded curve.

Component	Overall	1	2	3	4	5
Survival to Age 5 (A)	0.97	0.97	0.97	0.98	0.95	1.0000
Expected years of schooling	12.4	12.0	12.0	12.5	12.7	13.2
Harmonised test score	345	331	340	345	361	385
School (B)	0.85	0.69	0.71	0.74	0.78	0.85
Not stunted	0.79	0.72	0.79	0.82	0.86	0.88
Adult stunting rate	-	-	-	-	-	-
Health (C)	0.93	0.91	0.93	0.94	0.95	0.96
HCI (AxBxC)	0.66	0.60	0.64	0.67	0.71	0.81

Source: Authors own calculations.

As expected, the difference in the HCI between the income quintiles is now greater, with a 21 percentage point difference between income quintile 1 and 5 compared with 12 percentage points in Table 4.4. To contextualise this adjusted result: if South Africa had a national HCI estimate 21 percentage points greater than its current value (i.e. 0.62 instead of 0.41) it would rise to approximately 54 out of 157 countries from its current rank of 126 (World Bank, 2018b). This would place South Africa in approximately the top 40 percent of countries, instead of the current rank of the lowest 20 precent. Such is the advantage of income quintile 5 over quintile 1 households within South Africa.

[80] The descriptive statistics for household income per capita quintile 5 are: Observations (5761), mean (13512.55); Standard deviation (25101.16), minimum value 4600 and maximum 868507.80.

The SES-HCI is adjusted in Table 4.8 to include tertiary education disaggregated by income quintile. Equation 3.2 replaces equation 3.1 to estimate the schooling component, as explained above.

Table 4.8: SES-HCI, South Africa: Per income quintile, using a graded curve and mean education attainment at age 25.

Component	Overall	1	2	3	4	5
Survival to Age 5 (A)	0.97	0.97	0.97	0.98	0.95	1.0000
Expected years of schooling	11.3	10.6	10.7	11.2	11.6	12.5
Harmonised test score	345	331	340	345	361	385
School (B)*	0.58	0.54	0.55	0.57	0.61	0.69
Not stunted	0.79	0.72	0.79	0.82	0.86	0.88
Adult survival rate	-	-	-	-	-	-
Health (C)	0.93	0.91	0.93	0.94	0.95	0.96
HCI (AxBxC)	0.52	0.47	0.50	0.53	0.56	0.66

Notes: *Rate of return of 8% per additional year of education applied as per Kraay (2018).
Source: Authors own calculations.

The difference in HCI between income quintiles 1 and 5 is now 19 percentage points and the HCI measure is lower for all quintiles. Thus, although tertiary education was included because of its importance for labour market inequality, the uptake of tertiary education in South Africa is so small that the mean attainment of education is still only just over 12 years for quintile 5. If the SES-HCI was disaggregated further to income deciles, the difference would be greater, because the gap between mean years of attainment by age 25 jumps from the 10.22 and 13.7 for income deciles 1 and 10 respectively, compared with 10.68 and 12.50 for income quintiles 1 and 5. Calculating the SES-HCI by income decile would therefore more accurately reflect HC inequalities. The difficulty of disaggregating the data further into income deciles is due to the lower number of observations for each decile from the NIDS data, which raises questions of their validity. Furthermore, the impact of increasing rates of return to education in South Africa on SES-HCI estimates is recommended for future research given that authors including Van Broekhuizen (2011)[81], Branson and Leibbrandt (2013) and Lam *et al.* (2015) estimated high and increasing earnings returns from tertiary education. Despite the conclusion by Kraay (2018) that a consistent rate of return should be applied, given that income

[81] Van Broekhuizen (2011) used the wave 1 2008 NIDS dataset.

inequality originates in the labour market in South Africa, when disaggregating data to estimate a HCI for different SES indicators, it is recommended that future research applying different rates of return per education category, gender, race and geographical location should be conducted. The suggested research could furthermore be extended to estimating and applying economic rates of return to health outcomes specific to South Africa.

4.3.2 Disaggregation by geographical area, gender and race

In Table 4.10 the SES-HCI is disaggregated by geographical area, gender and race. First, the indices for schooling quality, based on the percentage of learners which attend the different school quintiles are presented in Table 4.9. This is done because of the importance of schooling in explaining SES-HCI differences. Schooling quality is better in urban than rural areas, is similar for females and males, but differs greatly between races, with Whites having an average HTS of 413 versus 340 for Africans. This difference is explained because 84.08% of Whites and only 7.30% of Africans attend quintile 5 schools. It should be noted, however, that approximately 90% of learners in South Africa are African, and 681 801 African learners were enrolled at quintile 5 schools as compared with 251 012 White learners (using NIDS weighted data). Thus, 57% of learners attending quintile 5 schools are African and 21% are White.

Table 4.9: Harmonised test score (HTS) by area, gender and race

School quintile	HTS	Percentage school quintile attendance by area, gender and race*							
		Area		Gender		Race			
		Urban	Rural	Male	Female	African	Coloured	Asian/Indian	White
1	315.33	12.86	32.18	21.16	22.34	23.80	6.00	4.16	1.96
2	319	15.85	35.64	25.35	24.56	26.84	13.31	2.08	2.14
3	338.33	30.06	25.9	29.31	26.97	30.03	16.61	1.16	6.57
4	392	21.73	4.34	13.01	14.45	12.02	37.03	29.31	5.26
5	424.67	19.5	1.95	11.16	11.69	7.30	27.05	63.29	84.08
TOTAL		100	100	100	100	100	100	100	100
Average score		361	328	345	346	340	378	407	413

Notes: *The calculation is based on individuals aged 4 to 17 enrolled in an educational institution.
Source: Authors own calculations.

In Table 4.10, Whites have an SES-HCI score of 0.64 compared with 0.58 for Asians/Indians, 0.54 for Coloureds and 0.50 for Africans. The schooling components and health components have 10 and five percentage point differences respectively between Africans and Whites. Coloureds have a six percentage point difference in the health component compared to Whites. It is important to note that stunting is the only indicator included in the health component and differences between races would likely increase if data were available to also include ASRs. According to Marandu (2011) the adult mortality rates of Africans, Coloureds, Asians/Indians and Whites were 0.613, 0.414, 0.311 and 0.260 respectively by the age of 60. Although the schooling component is the main influencer of HC difference in Table 4.10, survival and the health components are also important contributory factors. It is also important to note that the distribution of HC when measured by quintiles is more unequal according to race than according to income. The percentage point difference between income quintiles 1 and 5 is 12 percentage points (Table 4.4), whereas the difference between Africans and Whites is 14 percentage points.

Table 4.10: SES-HCI, South Africa: area, gender and race, using educational enrolment.

Component	Overall	Urban	Rural	Male	Female	African	Coloured	Asian/ Indian	White
Survival to Age 5 (A)	0.97	0.974	0.967	0.975	0.969	0.972	0.98	1.00	1.00
Expected years of schooling	12.4	12.5	12.2	12.1	12.6	12.3	12.5	12.1	13.2
Harmonised test score	345	361	328	345	346	340	378	407	413
School (B)	0.56	0.58	0.54	0.56	0.57	0.56	0.60	0.61	0.66
Not stunted	0.79	0.81	0.77	0.77	0.89	0.79	0.77	0.85	0.93
Adult stunting rate	-	-	-	-	-	-			
Health (C)	0.93	0.94	0.92	0.92	0.96	0.93	0.92	0.95	0.98
HCI (AxBxC)	0.51	0.53	0.49	0.50	0.53	0.50	0.54	0.58	0.64

Source: Authors own calculations.

The SES-HCI scores are also calculated using survival to age 5 (Table 4.11) (NDoH *et al.*, 2019) and HTSs graded on a curve (Table 4.12). The SADHS mortality rates result in the SES-HCI indicator for each measure being lower than the NIDS estimates. However, the inequalities of HC when calculated in this manner are not much different from the NIDS estimates.

Table 4.11: SES-HCI, South Africa: area, gender and race, using educational enrolment (using SADHS (2016) U5MRs).

Component	Overall	Urban	Rural	Male	Female	African	Coloured	Asian/ Indian*	White
Survival to Age 5 (A)	0.96	0.96	0.95	0.95	0.97	0.95	0.96	-	0.97
Expected years of schooling	12.4	12.5	12.2	12.1	12.6	12.3	12.5		13.2
Harmonised test score	345	361	328	345	346	340	378		413
School (B)	0.56	0.58	0.54	0.56	0.57	0.56	0.60		0.66
Not stunted	0.79	0.81	0.77	0.77	0.89	0.79	0.77		0.93
Adult stunting rate	-	-	-	-	-	-			
Health (C)	0.93	0.94	0.92	0.92	0.96	0.93	*0.92*		*0.98*
HCI (AxBxC)	0.50	0.52	0.48	0.49	0.53	0.49	0.53	-	0.62

Notes: *There were no results published by Victoria *et al.* (2020) for Asians and Indians thus excluded.

Source: Authors own calculations, NDoH *et al.* (2019) and Victoria *et al.* (2020).

As shown in Table 4.12, the graded curve education estimates accentuate the difference between the different SES indicators. The urban indicator is eight percentage points higher than the rural; female is five percentage points higher than male, and the gaps between Africans and other race groups and Whites increase even further. This is because schooling becomes the determining factor of differences between race groups. Adjusted for school quality, there is now a 19 percentage point difference in the schooling estimate between Africans and Whites.

Table 4.12: SES-HCI, South Africa: area, gender and race, using graded curve for HTS.

Component	Overall	Urban	Rural	Male	Female	African	Coloured	Asian/ Indian	White
Survival to Age 5 (A)	0.97	0.974	0.967	0.975	0.969	0.972	0.98	1.00	1.00
Expected years of schooling	12.4	12.5	12.2	12.1	12.6	12.3	12.5	12.1	13.2
Harmonised test score	345	361	328	345	346	340	378	407	413
School (B)	0.73	0.77	0.69	0.72	0.74	0.72	0.80	0.82	0.91
Not stunted	0.79	0.81	0.77	0.77	0.89	0.79	0.77	0.85	0.93
Adult stunting rate	-	-	-	-	-	-			
Health (C)	0.93	0.94	0.92	0.92	0.96	0.93	*0.92*	*0.95*	*0.98*
HCI (AxBxC)	0.66	0.70	0.62	0.64	0.69	0.65	0.72	0.78	0.89

Source: Authors own calculations.

4.3.3 SES-HCI disaggregated by school quintile

Because of the importance of schooling quality in HC differences, HC according to school quintile is shown for HTSs on a graded curve in Table 4.13. It is difficult to estimate under-5 mortality and stunting rates by school quintile since children have to be four or older to be in school. Thus, income quintiles are used as proxies for school quintile indicators because of the expectation that school quintile, survival and health indicators will be positively correlated to income quintiles, given the fact that higher quintile schools are fee-paying.

Table 4.13: SES-HCI, South Africa: Per school quintile.

Component	1	2	3	4	5	1	2	3	4	5
	Max HTS 625*					*Graded curve (max HTS 425)*				
Survival to Age 5 (A)	0.97	0.97	0.98	0.95	1.00	0.97	0.97	0.98	0.95	1.00
Expected years of schooling	13.1	12.9	13.1	13.2	13.6	13.1	12.9	13.1	13.2	13.6
Harmonised test score	316	319	338	392	425	316	319	338	392	425
School (B)	0.55	0.55	0.58	0.63	0.68	0.71	0.71	0.75	0.86	0.97
Not stunted	0.79	0.72	0.79	0.82	0.86	0.79	0.72	0.79	0.82	0.86
Adult stunting rate	-	-	-	-	-	-	-	-	-	-
Health (C)	0.91	0.93	0.94	0.95	0.96	0.91	0.93	0.94	0.95	0.96
HCI (AxBxC)	0.48	0.50	0.53	0.57	0.66	0.62	0.64	0.69	0.78	0.93

Source: Authors own calculations.

The SES-HCI difference between quintile 1 and quintile 5 schools is 18 percentage points, but 31 percentage points when applying the graded curve. The difference between quintile 1 and 2 schools is minimal, the largest difference being between quintiles 4 and 5. This is particularly evident for the graded curve where the difference between quintiles 4 and 5 is 15 percentage points. It is again evident that differences in the schooling component are driven by schooling quality rather than EYS.

4.4 Conclusions and Considerations

Based on household incomes, gender, geographical area, school quality and race, the estimated SES-HCI scores for South Africa highlight large inequalities in HC. The main determinant of this inequality is school quality. Children attending quintile 1 schools have a 18 percentage point lower SES-HCI score than those attending quintile 5 schools. The NIDS data suggest that only 11% of children attend quintile 5 schools. Thus, improving schooling quality, especially in quintile 1-4 schools, would reduce productivity differences and income inequality in South Africa.

These findings are supported by authors such as Leibbrandt *et al.* (2010), van der Berg (2010) and Spaull (2019). Emphasising the inequalities in education, Spaull (2019) found that 3% of high schools in South Africa produced more mathematics distinctions than the remaining 97% of high schools put together. It is evident that educational inequalities begin from a young age, with children from poorer households not obtaining basic skills and thus remaining perpetually disadvantaged when compared to children from wealthier households who attend better quality schools (Spaull, 2015). The cycle of poverty is thus perpetuated by the schooling system. As Taylor (2019) emphasized, it is clear that improved education and the quality thereof is the most likely way of improving the life-chances of children, regardless of race, language, gender, or geographical location. Furthermore, the improvement in educational quality is also required to drive the growth of a knowledge economy, as set out in the NDP 2030 (NPC, 2012), and has been identified by Asongu and Nwachukwu (2018) as required to ensure inclusive human development.

When applying a graded curve for HTSs, inequalities between the different SES indicators were exacerbated further. The large impact of using the graded curve on the HCI results illustrates how poorly South Africa performs globally in terms of schooling quality. Thus, in order to empower South Africans, improved learning and skills, rather than time spent at school, should be emphasised.

Gender inequality is also highlighted in the findings. Despite the SES-HCI estimates showing that women have higher HC than men, women have poorer employment prospects and earn less than men (StatsSA, 2019). Characteristics other than HC, such as additional penalties due to motherhood, thus negatively impact the labour market outcomes of women (Casale and Posel, 2005 and Magadla *et al.*, 2019).

These findings are aligned to the plans of the African Human Capital Project, which aims to increase productivity in Africa by 13% by 2023. The challenges highlighted as negatively affecting productivity in Africa as compared to other regions include girls and women who are born into a region experiencing the highest fertility rates, the percentage of girls who drop out of schools, maternal mortality rates and adolescent fertility rates. Increasing the learning-adjusted years of schooling and reducing adolescent birth rates by approximately 20% and 18% respectively have thus been identified as requirements to achieve the productivity improvements required (World Bank, 2019d).

Race was also identified as an important determinant of HC, which is unsurprising considering the Apartheid-era policies which practised racially biased expenditure on public services such as education. The discriminatory education system in the previous 'homelands' meant that Black schools not only received fewer resources than White schools but also had different education curricula, receiving a poorer quality of education (Timaeus *et al.*, 2013). While the post-Apartheid government has aimed to eliminate inequalities in public resource allocation to schools, education quality remains unequal and this inequality is identified in this study as perpetuating productivity gaps between population groups. As a result, earnings inequalities are particularly visible along racial lines in South Africa (StatsSA, 2019). The SES-HCI results confirm that labour market policies alone are insufficient to reduce racial income inequalities in South Africa. Improved education and health services are the key to reducing such inequalities.

While extending available research on HC distribution in South Africa by using the SES-HCI, this study has revealed several issues which need further consideration. Firstly, a SES-HCI which incorporates tertiary education would provide valuable information for estimating HC linked to the labour market. But for the results to be meaningful the data need~~s~~ to be disaggregated into income deciles rather than quintiles. Data availability is a challenge as was revealed in the NIDS data where only 26 deaths of under-5 children were evident from 870 births. Although the small number of observations in the NIDS data did not greatly impact the results in this study, disaggregating the data further, for example by income quintile to income decile, the more the estimates will become skewed and unreliable. In addition, further research accounting for higher and increasing earnings returns to higher levels of education and applying returns to education and health, specific to different subgroups in South Africa, should be explored. It is expected that greater inequalities would be identified.

Despite data limitations, estimating the SES-HCI for a number of indicators has provided important insights into the extent and causes of HC inequalities in South Africa. The SES-HCI estimates furthermore facilitate the monitoring of changes in HC gaps in South Africa.

CHAPTER 5

5 DECOMPOSING HOUSEHOLD INCOME INEQUALITY IN SOUTH AFRICA BY HUMAN CAPITAL DETERMINANTS.[82]

Abstract

Human capital (HC) has increasingly been identified as a driver of economic development, with the potential to reduce income inequality, which, in South Africa, originates in the labour market. HC is however a complex concept to measure. This study uses Fields' regression-based decomposition method to analyse the relationships between income inequality and HC. Fields' method allows the impact of several factors contributing to HC on the inequality of a measure of income. Data from the NIDS wave 1 (2008) and 5 (2017) is used. The findings suggest the dominance of education attainment in discussions of HC, is justified. The study confirms that increasing educational attainment, through improved school quality, is the remedy South Africa needs to reduce income inequality.

Keywords: human capital; education; inequality

[82] The article is under review in the South African Journal of Economics. The co-authors are Prof Keeton and Prof Rogan.

5.1 Introduction

Rank and Furrer (2016) stated that "persistently high and rising inequalities threaten social and political stability (Pieters, 2010) and the sustainability of democracies (Kapstein and Converse, 2008)". This is nowhere more relevant than in South Africa, a young democracy, with amongst the highest levels of wealth and income inequality in the world. The gini coefficients for wealth[83] and income in South Africa were 0.95 in 2017 (Orthofer, 2017) and 0.63 in 2014/2015 (World Bank, 2020) respectively, with no significant decreases in the inequality estimates over the 26 years since the end of apartheid (Chatterjee *et al.*, 2020). Furthermore, the Palma ratio[84] was estimated as 10.2 in 2015 and measured as the highest in the world. The World Bank (2012: xii) notes that the unjust reality in South Africa is that individuals' life outcomes are determined largely by whether "a person is born a boy or a girl, black or white, in a township or leafy suburb, to an educated and well-off parent or otherwise". Spaull (2020) concluded that predicting a child's future socioeconomic outcome (for example employment, income and education) in South Africa can be done with a certain level of accuracy by the age of seven. This means that many talented South African citizens are unlikely to reach their full potential, thus negatively impacting much needed social and economic development (Keeton, 2014).

Human capital (HC) has been identified as a driver of economic development with the potential to reduce income inequality (Evans, 2010 and Lee and Lee, 2018). HC is however a complex concept to measure and its importance for inequality is difficult to analyse. It is defined by the OECD (2001:18) as "the knowledge, skills, competencies and attributes embodied in individuals that facilitate the creation of personal, social and economic well-being". This paper measures the impact of several factors contributing to HC on the inequality of a measure of income in post-

[83] Wealth (also known as 'net worth') is calculated as the difference between the market value of all assets and liabilities (Orthofer, 2017) and in the National Income Dynamics Study (NIDS) household (HH) wealth is calculated by summing net financial wealth, net business equity, net real estate equity, value of vehicles, total value of pension/retirement annuities and livestock wealth (Chinhema *et al.*, 2016).

[84] Palma ratio is defined as the share of income received by the 10% of people with the highest disposable income divided by that received by the 40% of people with the lowest disposable income (OECD, 2020).

apartheid South Africa. This is done using Fields' regression-based decomposition method (Fields, 2003). The analysis is conducted across households (HH) and by population subgroups. The paper highlights the context of the causes of income inequality in South Africa with reference to the role of HC. In doing so, studies employing Fields' regression-based decomposition approach to investigate the determinants of inequality in developed and developing countries are first presented. Thereafter, the literature on measuring HC is discussed. Section 3 presents the data and methodology used to measure HC and decompose income inequality. Section 4 presents descriptive statistics related to HC and discusses the results from the decomposition of income inequality in South Africa. Section 5 concludes.

5.2 Context of income inequality in South Africa with reference to human capital

Studying inequality on a global scale, Piketty (2014) concluded that growing economic inequality in capitalist economies is primarily driven by income generated from wealth. Piketty (2014) explained that the rate of return on capital (*r*) is greater than economic growth (*g*) over the long-term, thus wealth-to-income ratios (the share of capital income) rise and growing wealth inequality becomes the source of economic and social instability. It was thus suggested that the growth rate of income from capital/wealth (terms used interchangeably) will almost always exceed that from labour, thus widening the inequality between the rich (owners of capital) and the poor (who own only their labour) (Piketty, 2014). The definition of what wealth is and thus how it should be measured has been expanded and debated (see Hodgson, 2015). Piketty (2014) classifies wealth as including "cash, bonds, shares and collateralizable assets such as buildings, land, machinery and intellectual property, but excludes social capital and non-slave HC" (Hodgson, 2015:1082).

A criticism of Piketty's (2014) study was that HC was not taken into account in measuring wealth/capital even though it is repeatedly recognized as an important driver of income. Given presumed higher productivity, it is therefore also an important element of HH wealth (Becker, 1962; Castelló and Domenech, 2002; Arrow *et al.,* 2012; Piketty, 2014; Keeton, 2014; Blume and Durlauf, 2015 and

Orthofer, 2017). Piketty (2014) justified the exclusion of HC from the analysis of economic inequality by stating that HC cannot be traded on a market or owned by another person and is thus difficult to measure. Hodgson (2015) supported Piketty's measure of capital, noting the importance of the collateralizable property of capital. Additional measurement difficulties of HC highlighted by Di Bartolo (1999), Costa (2006) and Lovaglio and Folloni (2011) are that, because it is statistically an unobserved latent construct, is multidimensional in nature and lacks a conclusive theoretical definition, there is a resultant lack of a prescribed observed variables.

However, observed links between HC and income and the links between HC and the labour market in South Africa mean that analysing poverty and inequality requires accurate means of measuring HC. Given the high levels of income inequality in South Africa and the lack of progress made in reducing inequalities since the end of apartheid, there is already a plethora of research on the factors causing income inequality in South Africa (Leibbrandt *et al.*, 2012; Van der Berg, 2014; Hundenborn *et al.*, 2016; Bassier and Woolard, 2018; and StatsSA, 2019). Hundenborn *et al.* (2016), analysed income from different sources, including labour, social grants, investments and remittances. The main cause of continued high income inequality levels in South Africa after 20 years of democracy was found to be differentials in the labour market. This finding is supported by several authors and organisations including Leibbrandt *et al.* (2012), Rani and Furrer (2016), Bassier and Woolard (2018) and the World Bank (2020). These authors found that wages have increased most for the top quintile of earners and for skilled workers. Van Broekhuizen (2011:97) found, using 2008 data, increasing returns to education in South Africa, and estimated the average marginal earnings in comparison to primary education attainment to be "Matric (35%), post-secondary diploma or certificate (83%), Bachelor's degree (115%) and post-graduate degree (180%)".

Over the period 1993 – 2008 growing wage differentials between skilled and unskilled labour (Pellicer and Ranchhod, 2012 and Leibbrandt *et al.*, 2010) were accompanied by rising unemployment, which exacerbated income inequality (Keeton, 2014). The unemployment rate in South Africa was 43.2% (if the statistics of discouraged workers are included) in the 1st Quarter of 2021. Of those unemployed, 52.4% had education levels below matric, in contrast to the 2.1% unemployment rate for university graduates (StatsSA, 2021). Thus, in addition to higher returns to education in terms of wages of the employed, the importance of education in improving employment prospects is a consistent finding in the post-apartheid literature (Pellicer and Ranchhod, 2012). South African research on the distribution of income at the end of the apartheid era focused on inter-racial income distribution (van der Berg and Louw, 2003). However, Van der Berg (2014) found that inequality within racial groups had increased. The most important factors recognized as driving intra-racial inequality are the unequal distribution of skills and inequality in educational outcomes. StatsSA's (2019) report on inequalities in South Africa illustrates that socioeconomic factors including race, gender and geographic region also play a major role in labour market outcomes.

In South Africa, research concerning HC and economic inequalities has focused on the impact of education and health policies on poverty and on the labour market (Lam, 1999; Liebbrandt *et al.*, 2010; van der Berg, 2010; and Pellicer and Ranchhod, 2012). Liebbrandt *et al.* (2010) noted that effective HC policies (education and health interventions) in South Africa are required to equip the labour force, particularly those disadvantaged by apartheid, with the skills that will be required for the labour force to become part of a productive workforce. Before poverty and inequality can be effectively addressed, such a productive workforce is necessary. This is relevant throughout Africa: the key to sustainable development requires workers to move upwards from low- to high-productivity jobs (Teal, 2021 and Kamanzi *et al.*, 2021). Van der Berg (2010) furthermore showed that reducing wage inequality requires improved HC development. Van der Berg (2002 & 2010) explored the impact of the quality of education versus the number of years of education attainment on outcomes in the labour market and concluded that in South Africa it was the quality of education that needed improvement. In studying HC accumulation and inequality traps in South Africa, Pellicer and Ranchhod (2012), assessed theoretical frameworks relating inequality and investment decisions in post-secondary education. They concluded that, rather than providing wider access, government policies should ensure the improvement of the quality of education. This is necessary in order to achieve the HC accumulation needed if inequality is to be reduced. Pellicer and Ranchhod (2012) furthermore highlighted that an inequality trap is evident in South Africa, where low incomes lead to low levels of skill accumulation, which perpetuates income inequality.

Spaull (2015) noted that inequalities in skills and educational outcomes in South Africa are driven by inequalities[85] in schooling. Spaull (2019) suggests it has become accepted that in South Africa there are two education systems, where no- or low-feepaying schools make up 75% of public schools. The remaining 20-25% of functional schools are fee-charging schools. (As an example of the differences in education quality in such a system, Spaull (2019) found that 3% of high schools in South Africa produced more mathematics distinctions than the remaining 97%. Furthermore, for every 100 learners who start school only seven graduate from university, with the main reason for dropouts being poor learning foundations (Spaull, 2015). Thus, Taylor (2019) stated that better-quality education is the most likely way of improving the life-chances of South African children regardless of race, language, gender or geographical location.

The health system in South Africa is also plagued with inequalities. This is illustrated by the unequal distribution of skilled healthcare professionals between public and private healthcare. Although 70% of doctors and specialists are employed in the private sector, only approximately 16% of the South African population can afford the high-quality services they provide (Makgoba, 2017).

It has thus been suggested that to reduce socioeconomic inequalities in South Africa, policies are required that focus on developing both the educational and health components of HC (Lam, 1999; Liebbrandt *et al.*, 2010; van der Berg, 2010; and Pellicer and Ranchhod, 2012). This is in line with microeconomic HC theory, which regards HC as a rational societal investment choice (Schultz, 1959 & 1961; Becker, 1962 & 1964; and Mincer, 1958 & 1974) which would influence "future real income through the embedding of resources in people" (Becker, 1962:9). The

[85] Public schools in South Africa are classified from quintile 1–5 (Q1-Q5) with Q1 being the poorest and quintile 5 being the wealthiest. No-fee paying schools are categorised as Q1-Q3, low-fee paying schools as Q4 and fee-paying schools as Q5 schools, which mainly consist of the formerly White 'Model C' schools (Spaull, 2019).

findings are also consistent with the social reproduction theory which posits that the education system, where HC is mainly developed, reinforces socioeconomic inequalities (Bourdieu and Passeron, 1977).

Fields (2003) provided a practical application of the regression-based decomposition method to explain the determinants of income inequality in the USA. The method has since been used in numerous studies (including Deng and Li, 2009; Manna and Regoli, 2012; Finnoff, 2015; Limanli, 2017; Tripathi, 2017 and Ayyash and Sek, 2020) to understand the determinants of income inequality in both developed and developing countries.

In each of the studies, education is identified as a common determinant of income inequality. Additional indicators are consequences of the history of the political economy in the various countries studied. For example, Deng and Li (2009), Finnoff (2015) and Limanli (2017), each studying different countries, identified historical regional development as an additional important determinant of income inequality. Deng and Li (2009) used a range of regression-based decomposition methods to decompose earnings inequalities within urban China between 1988 and 2002.

The different methods confirmed that provincial factors in urban China contributed the most to rising inequalities in income. The effects of work experience that were reduced over the period, as well as occupation and education, had increasing impacts. Thus, working in an urban region where jobs were predominantly skilled and associated with greater productivity has resulted in wage increases at a faster rate. Using the decomposition approach, Finnoff (2015) explored the determinants of poverty and income inequality in Rwanda in 2005, identifying education, region (urban/rural and province) and savings as having the largest impacts on income inequality. However, the determinant geographical area (urban versus rural areas) experienced the greatest decline in its contribution to income inequality between 2000 and 2005.

Limanli (2017) examined the determinants of income inequality in Turkey between 2006 and 2014, finding that education per capita and region were the greatest influencing factors and, furthermore, contributed most to changes in inequality. Ayyash and Sek (2020) explored the determinants of HH consumption expenditure inequality in Malaysia, confirming the importance of education, HH size and region as determining factors of inequality in that region.

Manna and Regoli (2012) and Tripathi (2017) included wealth as a determining impact of income inequality. Manna and Regoli (2012) compared HC and non-HC wealth as determinants of income inequality in Italy between 1998 and 2008, using the Fields and Shapley value regression-based decomposition approaches. Manna and Regoli (2012) found geographical area to have only marginal impacts on income inequality in Italy, while gender and HH wealth contributed the most. Tripathi (2017) applied Field's regression-based decomposition approach to the determinants of income inequality in India from 2004-2005 to 2011-2012, studying rural and urban subgroups. Increasing education levels and employment opportunities, together with proper land distribution policies, were identified as the most appropriate ways to reduce economic inequalities in India.

With each of these studies it is important to note that not more than 60% of income inequality was explained using the regression-based decomposition. Thus, the studies have suggested that the variables included in such a model should be explored within the context of the economy being studied as well as data availability (which Ayyash and Sek (2000) identified as a limiting factor). Different methods of measuring the concept HC are discussed next.

5.3 Measuring household human capital using an education-based approach

Le *et al.* (2005) identify three general approaches to measuring HC (Le *et al*, 2005). These are: an income-based measure, originally modeled by Farr (1853); a cost-based measure (Kendrick, 1976 and Eisner, 1989); and, lastly, various education-based methods (Barro and Lee, 1993 and Woßmann, 2003). Liu and Fraumeni (2021) categorised the HC measures into two broad approaches namely indicator-based and monetary measures. Towards providing researchers with comparable HC measures, Liu and Fraumeni (2021) discussed the methods and findings of six HC measures each of which are estimated across at least 130 countries. The monetary approaches discussed by Liu and Fraumeni (2021) are those developed to help government planning by including HC and nonhuman capital in a framework of national wealth accounting. The World Bank's report titled "Changing Wealth of Nations 2018: Building a Sustainable Future (CWON)" included a monetary value for HC for 141 countries over the period 1995 – 2014 by applying the Jorgenson-Fraumeni approach (Liu and Fraumeni, 2021). A second HC monetary measure is that presented in the biennial Inclusive Wealth Report (IWR) which included HC monetary values for 140 countries over the period 1990 – 2014. The HC measure was estimated by following a framework developed by Klenow and Rodriguez-Clare (1997) and Arrow *et al.* (2012, 2013).

Indicator-based measures discussed by Liu and Fraumeni (2021) include four largescale composite indexes for HC. The indexes include the World Bank's Human Capital Index (HCI), United Nation's Human Development Index (HDI)[86], the Institute for Health Metrics and Evaluation Human Capital Index (IHME HCI)[87] and the World Economic Forum's Global Human Capital Index[88] (WEF GHCI). The indexes use various macroeconomic education and health indicators as the two basic dimensions of HC towards estimating a counties' distance from optimal HC (the distance from maximum HC attainable)

Since the late twentieth century, an additional approach to measuring HC has evolved namely a latent-variable approach. The approach integrated the indicator- and income-based approaches by extending the measured latent HC into monetary terms (Dagum and Slottje, 2000). Dagum and Slottje (2000) defined HH HC as a latent variable formed by indicators including, in the words of Folloni and Vittadini (2010, 267), the "personal ability, social and home environments, investments in

---

[86] The HDI was developed to measure human development as a composite index and included in the first Human Development Report written by a team at the UNDP led by Mahbub ul Haq (UNDP, 1990). South Africa ranks 114th out of 189 countries (UNDP, 2020)

[87] The IHME HCI measures the "expected years lived from age 20 to 64 years and adjusted for educational attainment, learning, and functional health status" (Lim *et al.*, 2018: 1217). South Africa ranks 144th out of 195 countries.

[88] The WEF GHCI has been constructed since 2015 with the updated methodologies and most recent estimates presented for 130 economics for 2017 in The Global Human Capital Report 2017 (World Economic Forum, 2017). South Africa ranks 87th out of 130 countries.

the education of the household head and spouse". The importance of such indicators is reflected in HH income (Vittadini *et al.*, 2003). Despite the complex nature of HC, education attainment is frequently used as a proxy for HC, or even used as an interchangeable term (Finnoff, 2015). This is because education data is usually readily available (Barro and Lee, 2013) and because education is considered a fundamental component of HC.

Research using the education-based method of measuring HC is often focused on the aggregate level, comparing cross-country impacts of HC on income, development and economic growth. When critiquing the education-based method, Wößmann (2003) and Le *et al.* (2005) highlighted the different proxies originally used to specify HC, at a national level, namely: adult literacy rates, school enrolment ratios and the average years of schooling of the working age population. Wöβmann (2003) and Le *et al.* (2005) noted, however, that literacy rates are problematic measures as they capture only the first investment made into education, namely the ability to read. Enrolment ratios are also problematic, given that while HC is supposed to be measured as a stock, enrolment ratios include only children who are currently at school and so are not part of the labour force. Furthermore, in the South African context, Spaull (2015) has noted that although enrolment is near universal, education quality is very unequally distributed between schools. Using enrolment ratios is thus identified as a shortfall of the education-based and the World Bank's HCI approach to measuring HC.

Given the flaws of the two previously discussed proxies, average years of schooling, quantified by means of the highest educational attainment, has been used as a better measure of the national stock of HC (Wöβmann, 2003). Wöβmann (2003) however noted two main criticisms of years of schooling as an education-based HC proxy. The first criticism is the measure's failure to account for wage returns to schooling, given that one year of schooling does not directly raise labour market productivity and returns by the same amount, whether it is the first or 12th year of schooling (Wöβmann, 2003 and UN, 2016). Secondly, using average years of schooling to measure HC implies that all schooling systems are efficient and raise skills by the same amount (Wöβmann, 2003 and UN, 2016). A year of schooling should thus rather be weighted according to previously accumulated years as well as the quality of the education. In order for education to make a meaningful contribution to socioeconomic development, it needs to be of high quality and meet the needs of the skills demanded in the labour market. Meeting these requirements would ensure labour market outcomes reduce income inequality (Olaniyan and Okemakinde, 2008). A common response to overcome the first criticism has been to apply the Mincer earning function to calculate returns to education and to weight each year of schooling by market returns. Overcoming the second criticism has been more challenging and there is still no generally accepted solution to account for education quality differences (within and between countries) when measuring HC.

Among the three education measures described above, the educational attainment approach is the most widely used in South Africa when measuring HH HC. Several indicators have been used to proxy educational attainment at the HH level. These include: 1) Measuring the highest level of education attained by the head of the household (HHH) (van der Berg, 2002; Wanka, 2014; StatsSA, 2017; and World Bank, 2018a); 2) HH adult male and female highest educational attainment (Miller *et al.*, 2017); 3) the highest level of educational attainment by any individual in the HH (Mahoon *et al.*, 2003 and Thomson *et al.,* 2017); 4) the educational attainment by different HH members (van der Berg, 2002; Mahoon *et al.*, 2013; StatsSA, 2017, Hofmeyer, 2018 and World Bank, 2018a); and 5) a derived variable of aggregate HH education (Thomson *et al.,* 2017).

Rogan (2016) warned against using headship or a single reference person to classify HHs. This is particularly relevant to educational attainment in South Africa which consists predominantly of ‘extended' HH[89]s. In such a context, if educational attainment of the HHH was used to classify the educational attainment of the HH, the identification of the HHH as the oldest person in the HH would be a poor indicator. This is because the older generation were poorly educated during the Apartheid-era (The Economist, 2019). Thus, if educational attainment was proxied

---

[89] Hall *et al.* (2018) classify ‘extended' HHs as multi-generational or consisting of extended related family members living in one home.

by the members of the HH who were educated post-Apartheid, educational attainment would be estimated as relatively higher. The HH member/s educational attainment which would be the most accurate indicator to proxy HH HC thus needs to be context and data specific and, accordingly, different indicators are explored in this study. Moreover, the multidimensional nature of HH HC (OECD, 2001; Folloni and Vittadini, 2010; and World Bank, 2018a) and the large differences in education quality in South Africa (Spaull, 2019) are powerful arguments for moving beyond education attainment based on the HHH for measuring HC. The indicators used to analyse HH HC in an income generating model are presented in the method section. The data employed and methodological details of the regression-based decomposition approach are furthermore presented.

5.4 Method and data

5.4.1 Data and Sample

Primary data from the National Income Dynamics Study (NIDS) dataset are employed. The NIDS is a panel dataset compiled biennially by the South African Labour and Development Research Unit (SALDRU) based at the University of Cape Town (UCT). The initial (2008) wave of data was administered to a nationally representative sample[90] of approximately 28 000 individuals from 7 300 HHs, but grew with each wave as households expanded, to a total of 37 000 individuals and 13 719 HHs in wave 5. The most recent, Wave 5, was conducted in 2017 and released in 2018 (Brophy *et al.*, 2018). Waves 1 and 5 were used here to allow for an analysis of changes between 2008 and 2017. Additional reasons for using the NIDS dataset were, firstly, the very comprehensive nature of the socioeconomic data collected[91] and, secondly, that this research forms part of a larger study, using different approaches to measure HC and analyse HC distribution. Using the same dataset will make the results from the different studies comparable. Only HHs with

[90] The total numbers of individuals successfully interviewed in Wave 1 and Wave 5 were 26776 and 37 368 respectively (Brophy *et al.*, 2018).
[91] Examples of variables which will be used from the NIDS dataset include school administration data, grade repetition, school quintile, literacy, educational attainment, employment history, income, parents' education, stunting, mortality history, assets, debt and health.

data on the HHH were included in the study and thus 7 031 and 10 842 HHs comprise the samples in 2008 and 2017 respectively. Post-stratification weights were applied for both waves[92] of the NIDS dataset to allow for population estimates (SALDRU, 2017). For more detailed description of the data as well as survey methodology and weights, the interested reader should visit www.nids.uct.ac.za.

5.4.2 Analytical approach

The analytical approach included two steps. Firstly, exploratory data analysis was conducted by presenting the number of observations in the dataset, mean, standard deviation and correlation relationships between the variables used in the income generating model in Table 5.3. Secondly, income inequality is decomposed using Field's regression-based decomposition approach. The regression-based decomposition methodology was initially presented in the seminal work by Blinder (1973) and Oaxaca (1973). It gained further interest when Morduch and Sicular (2002) and Fields (2003) extended the decomposition of income inequality by income sources (Sharrocks, 1982) to income determinants (e.g. education, age, gender etc.), the latter being the focus of this study. Fields' (2003) decomposition method of income inequality starts by estimating a semi-*log* income generating model. The model developed for this study is an extension of the Mincerian earnings function, with the *log* of per capita income as the dependent variable. The exogenous variables are discussed in detail in subsection 4.3 below. The regression-based decomposition was conducted in the statistical program STATA using the package *ineqrbd* developed by Fiorio and Jenkins (2007) following the method proposed by Fields (2003).

[92] Weight variables w1_wgt (Wave 1) and w5_wgt (Wave 5) applied.

The study starts by considering the linear regression function:

$$logY_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots \beta_1 X_{i1} + \beta_k X_{ik} + \varepsilon_{i,} \qquad i = 1, 2, \ldots\ldots. n. \qquad (1)$$

Where $logY_i$ is the logarithm of net HH income per capita in ZAR for observation i and β_0, X_1, X_2 ,……X_{ik} are the exogenous explanatory variables (presented in Table 5.1); β_1, β_2,… β_k are the respective coefficients; ε_i is a vector of residual per observation (error term) and n is the sample size. Equation 1 is firstly estimated using the Ordinary Least Square (OLS) method. The results from the OLS are presented as part of the descriptive statistics of each variable in Table 5.4.

Thereafter, following the Fields (2003) decomposition approach, the relative factor inequality weight (s_k) for the kth predictor explanatory variable (share of the *log*-variance of income) is estimated by applying the following formula.

$$s_k = \frac{\widehat{\beta}_k \, Cov(x_k \,, lny_i)}{Var\ (\ lny_i)}, = 1, 2, \ldots, K \qquad (2)$$

Where $\hat{\beta}_k$ is the estimated coefficient for the kth predictor in equation (1); the outcome variable is $Var\ (\ lny_i)]$ and $Cov(x_k \,, lny_i)$ is the covariance between the kth predictor and the outcome variable.

The sign of the relative factor inequality weight (s_k) determines the direction of the factor's effect, and thus $s_k > 0$ or $s_k < 0$ indicates whether factor x_k is inequality increasing or decreasing, respectively. It is noted that the sum of all inequality weights should equal the R^2, as presented in equation 3, and thus the explained variation from the income generating model (equation 1).

$$\sum_{k=1}^{k} s_k = \frac{\sum_{k=1}^{k} \widehat{\beta}_k \, Cov(x_k \,, lny_i)}{Var\ (\ lny_i)} = \frac{Var\ (\ ln\hat{y}_i)}{Var\ (\ lny_i)} = R^2 \qquad (3)$$

Using Fields decomposition method (2003) the inequality decomposition may consider the error term $\varepsilon_{i,}$ as indicated in equation (1) (Fiorio and Jenkins, 2007), or the decomposition of predicted income may be applied and thus the residual in equation (1) is excluded. When the error term is considered, the unexplained inequality percentage allocated to it is symbolized by equation 4:

$$s_{\varepsilon} = 1 - R^2 \tag{4}$$

Fiorio and Jenkins (2007) identified that applying the command *ineqrbd* to Y_i (as illustrated in equation 1) versus predicted $\widehat{Y_i}$, affects the results given when decomposing Y_i, by the determining factors, the error term (ε_i) is also decomposed. The results for decomposing both predicted $\widehat{Y_i}$ and Y_i are presented in this study. A possible limitation of the Fields (2003) method is that it is limited to *log*-linear income generating functions. However, given that the income generating function is an extension of the Mincerian earnings function, this is not a limitation of this study.

5.4.3 Variables

From the South African literature, the factors related to HC that impact on income inequality have been identified as education, health, schooling quality, geographical area, parental well-being and employment status (Lam, 1999; Liebbrandt *et al.*, 2010; Van der Berg, 2010; Pellicer and Ranchhod, 2012; and StatsSA, 2019). The international studies, noted above, recognized education, region, wealth, gender and employment opportunities as factors determining income inequality (Ayyah and Sek, 2020). The variables selected for this study (presented in Table 5.1 below) are accordingly drawn from these studies as well as from the studies measuring HH HC as a multidimensional latent construct (Vittadini and Lovaglio, 2007; Dagum *et al.*, 2007; Lovaglio, 2008 and Lovaglio, 2010). Data availability is a further factor determining variable choice.

Table 5.1: Explanatory variables in the income generating model.

Human Capital Variables	*Exogenous factors impacting income*
Education-based indicators	
Average years of educational attainment of adults over 25 years and below 65 years of age (max 16 years per person)	Gender of HHH (Dummy 1 = male; 0 = other)
Health indicator	Race of HHH (Dummy 1 = White; 0 = otherwise)
Presence of member with poor self-perceived health status (Dummy 1 = No HH member has poor or fair health; 0 = Otherwise)	Employment status (No. of employed HH members)
Education quality indicator	Household geographical area (Dummy1 = urban; 0 = otherwise),
Presence of a working age adult who attended quintile 5 school (1 = One or more members have attended quintile 5 school; 0 = Otherwise)	Household size (continuous variable)
	HH age (mean age of working age HH members)

The analysis was conducted across HHs based upon an assumption that resources are shared amongst HH members and thus that the wellbeing of the individual depends on that of the HH (Pieters, 2011). The study was also conducted for racial subgroups (White and Black[93]), geographical area (urban and rural) and gender. This is to identify the inequalities within groups caused by factors which, in the South African context particularly, impact labour market outcomes (Leibbrandt *et al.*, 2010 and van Broekhuizen, 2011). Given that the analysis is focused at a HH level, it is important to note that the NIDS allows HH members to self-define the HHH. NIDS (2020) stated that when identifying the HHH "No guidance is given that the HHH must be the eldest, highest earner or of a specific gender". Thus, it appears the HHH variable in the NIDS dataset is a reference person (Budlender, 2003) rather than the main decision-makers or an economic contributor.

In view of the context of high levels of unemployment in South Africa and potential sample selection bias, two conceptual models were evaluated. The two models had the following differences: 1) in Model 1 the endogenous variable is the *log* of HH income per capita[94] and a control variable accounting for the number of employed

[93] African, Coloured, Asian and Indian ethnic groups are classified as Black in accordance with the Broad-Based Black Economic Empowerment (B-BBEE) Act 53 of 2003.

[94] Household income per capita is calculated based on the net income, which includes net labour, wealth, government transfers and remittances as classified by the NIDS. All these incomes affect wellbeing and income inequality and are thus included. The individual incomes in the HH were aggregated and then divided by the HH size. Each HH per capita observation was included once per HH.

HH members is included; 2) in Model 2 the endogenous variable is *log* of HH labour income per capita, but there is no control variable for employment. The rationale for Model 1 is that, given the high unemployment rate in South Africa, it allows for the inclusion of the unemployed in the sample group. However, it is expected that the contribution of HC determinants to income inequality would be under-estimated, because the impact of factors impacting both labour force participation and earnings will not be estimated in isolation. Model 2 excludes the unemployed and measures the contribution of HC determinants directly on HH labour income per capita. However, people who work are not a random sample from the population, thus estimating HC determinants only for a sample of individuals who are employed is subject to sample bias. To account for such biases in earnings models, researchers include Heckman selection corrections (Branson and Leibbrandt, 2012). However, the Fields regression-based decomposition method does not include the option of selection model correction. Therefore, to confirm the results, two models were analysed.

The variables related to HC and additional determinants of income in the study are expanded on next.

5.4.3.1 Human Capital variables

Studies which have employed Fields' (2003) decomposition method in a HH income generating model tend to employ the educational attainment of the HHH (Pandey, 2015; Finnoff, 2015 and Ayyash and Sek, 2020). However, considering the warning by Rogan (2016) against using headship or a single reference person to classify HHs in South Africa, the education attainment indicator in this study is the average years of education for all HH adults over the age of 25 and less than 65 years of age. The NIDS education attainment data was accordingly coded as per Branson and Leibbrandt (2013).

To account for schooling quality, a dummy variable was included to identify the presence of a member who has attended a quintile 5 school (Spaull, 2019). Quintile

5 schools are identified as providing significantly higher quality education and are thus used as a reference group (as suggested by Spaull and Kotze, 2015).

Health is measured using a dummy variable, which indicates the presence of any HH member who has poor self-perceived health. Since it is self-perceived and self-reported and thus subject to bias, this variable has a potential limitation. Nwosu (2015) identified the arguments for and against using subjective self-perceived health status variables to proxy true health. The arguments in support are the ability to capture various dimensions of health (Bound, 1991) and future mortality (DeSalvo *et al.*, 2005; and Ardington and Gasealahwe, 2012). However, non-random measurement errors of self-perceived health indicators highlight potential bias (Kreider, 1999). The conclusion reached by Nwosu (2015) was that, in capturing health and labour market participation relationships (Bound, 1991), self-perceived health responses may be more effective than reported individual illness or symptom responses. Furthermore, Ardington and Gasealahwe (2012) identify that, despite some measurement error, the health variables in the NIDS dataset provide for the analysis of health and socioeconomic status relationships.

5.4.3.2 Additional determinants of income

The remaining variables, namely employment status, age, race, gender, geographical region and HH size are included as factors that have been found to correlate with labour market income (StatsSA, 2019). The number of employed HH members represents the employment status of the HH and is therefore restricted to Model 1. The average age of working age adults is included as a control variable, as per the Mincerian earnings function (Manna and Regoli, 2012 and Branson and Leibbrandt, 2013). The HHH was used to proxy the race, gender and geographical area of the HH.

5.5 Results

5.5.1 Descriptive statistics

Table 5.2 below presents the HH income per capita[95] distribution for 2008 and 2017 estimated by calculating gini coefficients for the South African population, as well as by race, gender and geographical area of the HHH. The results show that the only indicator for which the gini coefficient estimate increased slightly over the 9-year period is that for White HHs. This group has the lowest gini coefficient and thus the lowest inequality in distribution of HH income per capita in both periods. Inequality is higher in urban than rural HHs in both periods.

Table 5.2: Household income per capita gini coefficient estimates.

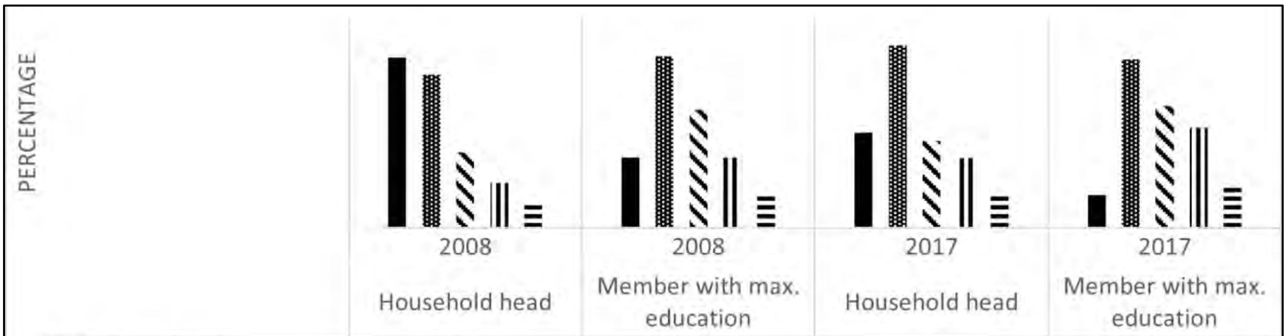
	Gini co-efficient	
	2008	2017
Household income per capita	0.67	0.65
Subgroup:		
Black household head	0.62	0.61
White household head	0.47	0.48
Male household head	0.65	0.63
Female household head	0.66	0.65
Urban households	0.64	0.64
Rural households	0.61	0.59

Source: NIDS wave 1 and wave 5; author's calculations. Post-stratified weights applied.

The distribution of the educational attainment of the HHH and HH member with the maximum amount of education attainment is presented for 2008 and 2017 in Figure 5.1 below. It is evident that using different HH members' educational attainments to estimate HC will affect the results. Using the HH member with the maximum amount of education rather than the education of the HHH will reduce the percentage of HHs with lower levels and increase the percentage with higher levels of educational attainment. The largest difference occurs for the lowest measure of education where there are more than double the number of observations of less than grade 7 for the HHH, when compared with the most educated member of the HH for

[95] The HH income inequality per capita variable is represented by one observation per HH, which is identified using the HHH.

both 2008 and 2017. The educational attainment gini coefficients were estimated as 0.29 for the HHH and 0.24 for the most educated HH member in 2008. For 2017, the respective gini coefficient estimates are 0.26 and 0.21. The gini coefficient has thus decreased for both categories. Furthermore, it is evident that overall educational attainment has improved over the 9 years analysed. For example, 15.10% of the most educated HH members were found to have less than Grade 7 in 2008, compared to 6.97% in 2017. The share amongst the most educated HH members with some tertiary education or Bachelor degrees rose from 22.66% to 30.77%.



	2008 Household head	2008 Member with max. education	2017 Household head	2017 Member with max. education
Grade 7 or less	36,62	15,10	20,38	6,97
Secondary no Matric	32,85	36,86	39,10	36,18
Matric or equivalent	16,21	25,38	18,56	26,09
Some tertiary	9,53	15,09	14,97	21,50
Bachelor Degree or More	4,79	7,57	6,99	9,27

Figure 5.1: Educational attainment of household head and the household member with the maximum amount of education, 2008-2017.
Source: NIDS wave 1 and wave 5; author's calculations. Post-stratified weights applied.

In Table 5.3 the descriptive statistics for each of the variables included in the income generating models are presented. For the continuous variables, the mean and standard deviation are presented, whereas for the dummy variables, the percentage observations for the control (0) and the dummy (1) group are presented.

Table 5.3: Descriptive statistics of NIDS wave 1 & wave 5 datasets: 2008 & 2017.

Variable	Observations		Mean (continuous)/Percentage (Categorical or			
	2008	2017	2008	2017	2008	2017
					Std dev.	
HH income per capita	7031	10841	2482.77	6165.54	4973.18	15933.18
HH labour income per	4569	6506	2442.18	5193.62	4735.09	8867.35
Average education	6215	9484	8.91	10.48	3.87	3.01
Quintile 5 school attended	5147	9006	0= 81.10% 1=18.90%	0=84.10 1=15.90		
Health indicators						
Self-perceived health status	6989	10634	0=28.64% 1=71.36%	0=16.65% 1=83.35%		
Additional exogenous income indicators						
Gender of HHH	7031	10841	0=38.53% 1=61.47%	0=47.04% 1=52.96%		
Race of HHH	6876	10737	0=86.75% 1=13.25%	0=90.18% 1=9.82%		
Employment status	7031	10841	0.86	0.97	0.81	0.84
Household geographical area	7031	10841	0=33.21% 1=66.79%	0=29.56% 1=70.44%		
Household size	7031	10841	3.51	3.03	2.51	2.38
Mean age of working age adults	6725	10395	35.06	35.03	9.88	9.30

Source: NIDS wave1 and wave 5; author's calculations. Post-stratified sampling weights applied.

As noted previously, there are fewer observations in the 2008 dataset than in 2017. With regard to HH income per capita, the standard deviation in 2017 of R15 933.18 is more than double the mean, which is indicative of the high levels of income inequality in South Africa. The data shows that the mean years of education attainment per household of adults aged 25 to 65 has increased over the 9-year period, from an average of 8.91 to 10.48 years. Both means are however less than 12 years, which would have been the case if they had completed secondary schooling. Most HHs do not have even one individual who has attended a quintile 5 school. Only 18.90% of HHs had a member who has attended a quintile 5 school in 2008, and this fell to 15.90% in 2017. In both years studied the majority of HHs were male headed. However, this has reduced, with female headship increasing from 38.53% in 2008 to 47.04% in 2017. 86.75% HHs had Black HH heads in 2008, and this increased to 90.18% in 2017. With regards to employment status, fewer than 1 person per HH is employed on average, which is indicative of the high unemployment rate in South Africa. The average household size for both years is between 3 and 4 individuals and the mean age of working-age adults has remained constant at approximately 35 years of age. The descriptive statistics are all consistent with expectations in the South African context. Given that post-stratification sampling weights are applied to the data, the samples can be interpreted as representative of the South African population.

5.5.2 Income generating model

The results of both income generating models, from which the regression-based decomposition is developed, are shown in Table 5.4 below. The results from the income generating model showed results which were consistent with expectations and fit the data well. Considering the human capital indicators, education has the biggest impact and each indicator has a greater impact when restricted to the labour market. Furthermore, race and gender are largely correlated with income.

The R-squared for Model 1 is 0.6239 in 2008 and 0.5480 in 2017. This translates to 62.39% and 54.80% of the measured income inequality being explained by the regression-based decomposition. None of the international studies discussed had an R-squared value of more than 0.6 and thus the regression model for both years in this study compares favourably with what has been found elsewhere. The residual estimates of the regression-based decomposition are thus on the lower end of the previous studies discussed. Model 2 likewise confirms acceptable residuals, with R-squared values of 0.5643 in 2008 and 0.4776 in 2017. However, when interpreting the factor contributions of independent variables to income inequality, it is important to note that Model 2 (2017) will only explain 47.76% of the inequality.

Table 5.4: Slope coefficients from the income generating models.

Variable in Model	NIDS Wave 1 (2008)		NIDS Wave 5 (2017)	
MODEL 1				
Dependant variable: *log* HH income per capita	β in OLS	Std error	β in OLS	Std error
Human capital factors				
Education-based indicators				
Average education	0.1293	0.0041	0.1420	0.0036
Quintile 5 school attended	0.1328	0.0378	0.2147	0.0285
Health indicators				
Self-perceived Health status	0.0140*	0.0285	0.0989	0.0243
Additional exogenous income indicators				
Gender of HHH	0.3032	0.0261	0.2370	0.0186
Race of HHH	0.9455	0.0496	0.8302	0.0431
Employment status	0.4112	0.0149	0.3896	0.0112
Household geographical area	0.2535	0.0283	0.0919	0.0204
Household size	-0.1354	0.0052	-0.1803	0.0041
Mean age of working age	0.0232	0.0016	0.0165	0.0012
R2/residual	0.6239		0.5480	
MODEL 2				
Dependant variable: *log* HH wage per capita				
Human capital factors				
Education-based indicators				
Average education	0.1844	0.0062	0.2009	0.0055
Quintile 5 school attended	0.1578	0.0520	0.3245	0.0423
Health indicators				
Self-perceived health status	0.2768	0.0438	0.2270	0.0389
Additional exogenous income indicators				
Gender of HHH	0.5427	0.0391	0.4208	0.0280
Race of HHH	0.6905	0.0684	0.6170	0.0622
Household geographical area	0.2899	0.0434	0.1729	0.0313
Household size	-0.1445	0.0075	-0.1713	0.0060
Mean age of working age	0.0279	0.0025	0.0140	0.0018
R2/residual	0.5643	--	0.4776	--

Notes: *Not statistically significant ($p > 0.05$).

Source: NIDS wave 1 and wave 5; author's calculations. Post-stratified weights applied.

With regards to the average education attainment of the HH, the results from Model 1 in 2008 show that an increase of one year of education attainment increased income by 12.93%, while controlling for the remaining variables. The impact of education attainment on employment is however not captured. An additional person employed in the HH is estimated to increase income by 41.12%. By restricting the model to only those employed (Model 2), the impact of human capital variables increased. A 1-year increase in education attainment increases income by 18.44%. Given the results are indicative of a 1-year increase in education it is evident that, for example, increasing education by 5 years will increase labour income per capita by nearly 100%. Furthermore, the marginal increase is expected to be higher for higher years of education, as identified by van Broekhuizen (2011) and Branson and Leibbrandt (2013).

The returns to education have increased over the study period for both models. Model 1 suggests a return per year of education of 14.20% and Model 2 a return of 20.09%. In Model 2, a HH with a member who attended a quintile 5 school is expected to be associated with a household per capita income 13.28% greater in 2008 and 21.47% greater in 2017 *ceteris paribus*. Similar to average years of education, this is greater for Model 2, which measures returns specific to the labour market.

Health, although showing statistically insignificant results in 2008 in Model 1, is noted as having a positive relationship with income in both models. However, in 2017, its impact, while significant, is the lowest of the HC variables in Model 1. Income per capita for a HH without a member with poor health is only 9.89% greater than HH with a member in poor health. It should however be cautioned that this finding could be a result of the variable used to proxy health being self-reported. When restricting the impact of health to labour earning in Model 2, the impact of health on HH labour income per capita in 2008 is now significant. Thus, health appears to impact HH income through labour market earnings.

The schooling quality indicator (the presence of a HH member having attended a quintile 5 school) also shows greater impacts on income when restricted to the labour market (Model 2) and its importance increased substantially between 2008 and 2017. In Model 2 HH labour income per capita increases by 15.78% in 2008 and 32.45% in 2017 for HH having a person who attends a quintile 5 school, while controlling for the other variables in the model. However, when interpreting school quintile results, it should be cautioned that school quintile may be an indicator of wealth, with only members of wealthier HH attending quintile 5 schools (Spaull, 2015).

With regards to the additional exogenous indicators of income, unsurprisingly, race plays an important role in income generation. White-headed HHs had per capita incomes 94.55% higher in 2008 and 83.02% higher in 2017 than HHs headed by other race groups. In Model 1, employment status is the next highest contributor, with HH income per capita increasing by 41.12% in 2008 and 38.96% in 2017 per person employed in the HH. The gender of the HHH also has significant impacts on HH income per capita in both Model 1 and 2. This has however decreased over the study period with the increase in income caused by having a male HH head (compared to female) falling from 30.32% in 2008 to 23.70% in 2017 for Model 1 and from 54.27% in 2008 to 42.08% in 2017 for Model 2. Geographical area and household size also impact on HH income per capita. The average age of the HH working age members however has a negligible impact in both models.

Having identified the returns to income per capita of the different exogenous variables in the income generating models, the factor contributions of each variable to income inequality estimated are presented in Table 5.5 below.

5.5.3 Regression-based inequality decomposition results

The contributions of HH determinants to income inequality are firstly presented for both models for 2008 and 2017 at a population level (Table 5.5). The results for the subgroups race, gender and geographical area are shown in Tables 5.6 and 5.7. As noted by Fiorio and Jenkins (2007)[96] the decomposition results can be presented for Y_i (equation 1 above) as well as predicted $\widehat{Y}_i$. The results for decomposing both predicted $\widehat{Y}_i$ and Y_i are presented when analysing the population (Table 5.5). However, for ease of comparison between the subgroups (Table 5.6 and 5.7), predicted $\widehat{Y}_i$ is decomposed and thus the error term (ε_i) is excluded from the decomposition.

---

[96] The fields option is used to decompose predicted $\widehat{Y}_i$ using the *ineqrbd* command in Stata (Fiorio and Jenkins, 2007).

The findings below highlight that education attainment is consistently a dominant determinant of income inequality. While household size and race are also important determinants, household size is noted as having an increasing impact whereas the contribution of race has decreased over the 9-year time period. The health and school quality indicators are highlighted as small contributors to inequality, despite the positive relationship between the indicators and income in the income generating model. Furthermore, the contributions of health and schooling quality are suggested as factors determining education attainment and labour market participation and thus indirectly impacting income inequality. Employment status had a lower contributory impact on income inequality than expected, highlighting the large income inequalities which occur in the job market.

At the subgroup level, the results below indicate that education attainment continues to be dominant, with household size making up most of the difference where differences between subgroups exist. Employment status was recognized as also contributing to the differences between subgroups in Model 1. The findings at the population level were mostly confirmed at the subgroup level.

5.5.3.1 Household income per capita for the total population

The inequality weight for each factor presented in Table 5.5 below is estimated as a function of the income generating model employed, the covariance between the *log* income and the factor studied, as well as the variance of the *log* income. The factor weights discussed below are for the decomposition of predicted $\widehat{Y_i}$.

Table 5.5: Factor contributions to inequality using Fields' Regression-Based Decomposition: 2008 - 2017.

	NIDS Wave 1 (2008)		NIDS Wave 5 (2017)	
	lnY_i	Predicted $ln\widehat{Y_i}$	lnY_i	Predicted $ln\widehat{Y_i}$
Dependant variable: *log* HH income per capita	Factor inequality weight sjx100	Percentage contribution net of residual	Factor inequality weight sjx100	Percentage contribution net of residual
Human capital factors				
Education-based indicators				
Average education	19.44	31.16	16.55	30.19
Quintile 5 school attended	1.45	2.33	1.82	3.32
Health indicators				
Self-perceived Health status	0.11*	0.18*	0.52	0.94
Additional exogenous				
Gender of HHH	3.49	5.60	2.84	5.18
Race of HHH	11.05	17.71	6.34	11.57
Employment status	7.80	12.50	6.66	12.15
Household geographical area	3.36	5.39	1.00	1.83
Household size	11.42	18.31	16.93	30.89
Mean age of working age	4.26	6.83	2.15	3.93
R2/residual	37.61	--	45.20	
MODEL 2				
Dependant variable: *log* HH				
Human capital factors				
Education-based indicators				
Average education	22.95	40.67	20.10	42.08
Quintile 5 school attended	1.29	2.29	2.07	4.33
Health indicators				
Self-perceived Health status	2.58	4.57	1.18	2.46
Additional exogenous				
Gender of HHH	5.01	8.88	4.18	8.76
Race of HHH	6.38	11.31	3.52	7.38
Household geographical area	2.64	4.68	1.33	2.78
Household size	11.67	20.68	14.34	30.02
Mean age of working age	3.91	6.92	1.05	2.19
R2/residual	43.57	--	52.24	--

Notes: *Not statistically significant (p > 0.05).

Source: NIDS wave 1 and wave 5; author's calculations. Post-stratified weights applied.

The results for both models for 2018 show (in descending order) that education, HH size and the race of the HHH contribute the most to income inequality in South Africa. The contribution of race falls from 17.71% in 2008 to 11.57% in 2017 for Model 1 and from 11. 31% to 7.38% for Model 2. Thus, causes of intra-racial inequality are interrogated and will be highlighted in the subgroup decomposition.

In 2017, race, as the third most important factor contributing to income inequality, is replaced by employment for Model 1 and by gender of the HHH for Model 2. The contribution of employment status in Model 1, over both years, is surprisingly low: 12.5% in 2008 and 12.15% in 2017 of predicted $\widehat{Y_i}$. It thus appears that HH income per capita inequality is driven more by the income earned by unskilled compared to skilled workers. Thus, inequality lies in the labour market (as suggested by Leibbrandt *et al.* [2012]) instead of in the number of employed HH members. This confirms that transferring workers to skilled and better paying jobs is the key to sustainable development throughout Africa (Teal, 2021 and Kamanzi *et al.*, 2021).

Focusing on the youth in Tanzania, Kamazi *et al.* (2021) found that this movement is driven by education and household factors, which provide different implications for males and females (Kamazi *et al.* 2021). The progression of girls towards obtaining high-quality jobs was restricted by their having to assume responsibilities of caring. Furthermore, an intergenerational transmission of skills was evident where girls whose fathers had higher levels of education had additional advantages.

Years of education is a HC indicator which has a dominant impact on income inequality in both models. For example, for Model 1, the contribution of education attainment is 31.16% in 2008 and 30.19% in 2017. Thus, increasing educational attainment is confirmed as the most likely way of reducing income inequality. This is particularly evident in Model 2, which represents the labour market and supports the conclusion that this is where income in South Africa originates (Leibbrandt *et al.*, 2012).

The contributions of the remaining human capital determinants, health and schooling quality, however, are found to have minor contributing impacts. The combined contribution of health[97] and school quintile is only 2.51% in 2008 and 4.26% in 2017. Thus, over and above the controls for labour market attachment, at a population level, it appears that the dominance in previous studies of using education attainment to measure the impact of HC on income inequality is justified.

[97] The indicator for good health was however not statistically significant.

It should however be noted that the significant income earnings return to good health (as presented in Table 5.4), do not appear to be captured in the decomposition analysis.

With regards to the schooling quality indicator, although the importance for inequality of working age adults having attended a quintile 5 school is low, the effects of the quality school attended could be a factor determining higher educational attainment. Spaull and Kotze (2015) noted that the learning trajectory of a learner in South Africa is strongly influenced by the quality of school attended. Using the Wave 5 NIDS data it is estimated that only 3.91% of working age adults over the age of 25, who attended quintile 1-4 schools obtained a Bachelor degree, in comparison to 16.25% who attended quintile 5 schools.

The determinant whose contribution to income inequality has increased most over the period is HH size - from 18.31% in 2008 to 30.89% in 2017 for Model 1. Thus, per-person welfare is increasingly determined by HH welfare, because resources are shared amongst household members. Next the contribution of the HC factors on income inequality are presented by gender, race and geographical area sub-groups.

5.5.3.2 Household income per capita by gender, race and geographical area

Average years of education is yet again the dominant HC factor impacting on income inequality by subgroup. For example, for White- and Black-headed HHs, in 2017 and Model 1, average years of education determines 44% and 32.53% of income inequality, respectively. In comparison school quintile and health contribute a combined value of only 1.27% for White-headed HHs and 2.63 for Black-headed HHs. Compared to Model 1, Model 2 furthermore indicates that educational attainment is a greater contributor to income inequality for all years and across all subgroups. The importance of education within the labour market where income inequality is generated is thus confirmed. Some results for self-perceived health as well as school quintile are not statistically significant, as presented in Table 5.6 below.

The dominant contribution of education attainment as an indicator of HC to income inequality thus justifies using education attainment as a proxy for HC in South Africa and as a determinant of income inequality, particularly in the labour market. However, the contributions of school quality and health to education attainment and labour market participation are potentially where the impact on income inequality occurs. This is identified as an area for future research. An example of existing research is that of Nwosu (2015) who, using the NIDS dataset in a cross-sectional analysis, found statistically significant impacts of health (using different measures) on labour force participation of between 20% and 33%. Furthermore, the results of the income generating model in Table 5.4, reveal that (having controlled for the remaining variables in the model) both health and education quality clearly influence earnings in the labour market. When studying the variance in the *log* wages, however, education attainment becomes the dominant determinant.

Table 5.6: Contribution of human capital determinants of income inequality, by subgroup.

	HH per capita total income (Model 1)		HH per capita labour income (Model 2)	
Sub-group	**Predicted $ln\widehat{Y_i}$			
Human capital factors	2008	2017	2008	2017
Average Education				
Race				
White	68.06 (0.3370)	44.00 (0.4384)	54.91 (0.2742)	52.18 (0.3717)
Black	34.01 (0.5232)	32.53 (0.4821)	43.76 (0.4867)	44.81 (0.4290)
Region				
Urban	34.38 (0.5994)	35.10 (0.5325)	42.45 (0.5408)	46.96 (0.4631)
Rural	28.08 (0.4778)	21.89 (0.4954)	40.04 (0.4489)	34.80 (0.4268)
Gender				
Male	39.50 (0.6086)	37.62 (0.4950)	46.67 (0.5594)	46.66 (0.4126)
Female	26.53 (0.5629)	30.01 (0.5352)	45.20 (0.4586)	51.64 (0.4691)
Quintile 5 school				
Race				
White	0.91*	0.02	0.95*	1.30*
Black	1.57	1.53	1.13	1.70
Region				
Urban	1.68*	2.91	2.55	4.00
Rural	1.11	3.29	0.50*	3.86
Gender				
Male	2.84	2.52	3.02	4.75
Female	1.62*	5.22	1.95*	5.32
Self-perceived health				
Race				
White	-0.35*	1.25	1.33*	5.81
Black	0.39*	1.10	6.21	2.70
Region				
Urban	0.86*	1.17	5.35	2.75
Rural	-1.43	0.29*	3.55	1.89
Gender				
Male	0.80*	1.37	4.96	2.80
Female	-0.61*	0.62	4.57	2.38

Notes: * Not statistically significant ($p > 0.05$). **R-squared per sub-group model is presented in brackets after the factor contribution of average years of education attainment.

Source: NIDS wave 1 and wave 5; author's calculations. Post-stratified weights applied.

Across the subgroups, for both models and time periods, average years of education has a greater determining influence on income inequality for Whites than for Blacks, as well as in urban compared to rural areas, and for males compared to females[98]. One reason for this could be the relationship between education attainment resulting in higher productivity jobs available to the White, urban and male-headed

[98] The only exception is for gender for Model 2 in 2017.

households (Kamanzi *et al.,* 2021). Furthermore, investigating the determinants between the different subgroups reveals that HH size and the number of HH members employed (relevant to Model 1) are the factors which make up most of the difference in the contribution of education attainment for both the years studied. For example, for Model 2, household size is found to contribute 38.16% and 23.18% to labour market income inequality for Black and White HHs respectively in 2017. Thus, the burden of sharing resources amongst more household members is evident for all race groups, but is greater for the Black subgroup. The same is evident for urban compared to rural and male- compared to female-headed households. The relationship between HH size and income per capita is furthermore bidirectional, given that HHs with fewer resources tend also to be larger. In 2017, 2%, 3.21% and 66.95% of HHs with only one HH member were in income quintile 1, 2 and 5, respectively. In contrast, 35.23%, 30.82% and 3% of HHs with 10 HH members were in income quintile 1, 2 and 5, respectively. To investigate the contributions of the exogenous indicators on income inequality further, Table 5.7[99] presents the results per subgroup.

[99] The average age of the working age members of the HH was not presented in Table 5.7 given the negligible contributions noted in Table 5.5 above.

Table 5.7: Contribution of HH characteristics to income inequality, by subgroup.

	HH per capita total income		HH per capita labour income	
	**Predicted $ln\widehat{Y}_i$			
	2008	2017	2008	2017
HOUSEHOLD SIZE				
Race				
White	17.17 (0.3370)	37.85(0.4384)	19.06 (0.2742)	28.13(0.3717)
Black	24.79 (0.5232)	39.53(0.4821)	26.08(0.4867)	36.18(0.4290)
Region				
Urban	13.14 (0.5994)	27.03(0.5325)	17.28(0.5408)	26.67(0.4631)
Rural	39.82(0.4778)	44.19(0.4954)	37.90(0.4489)	41.31(0.4268)
Gender				
Male	18.98(0.6086)	30.23(0.4950)	23.42(0.5594)	32.22(0.4126)
Female	18.41(0.5629)	29.34(0.5352)	19.13(0.4586)	26.57(0.4691)
RACE				
Region				
Urban	20.84	13.81	14.44	9.72
Rural	2.58	1.24	0.30	0.41
Gender				
Male	16.27	16.25	11.67	10.30
Female	24.13	8.92	14.65	6.49
GENDER				
Race				
White	6.10	5.50	18.55	5.40
Black	7.67	6.16	10.74	10.32
Region				
Urban	5.93	3.62	9.84	7.92
Rural	7.20	10.66	10.56	13.75
GEOGRAPHICAL AREA				
Race				
White	0.70	0.99	5.43	2.22
Black	6.23	1.96	4.63	2.80
Gender				
Male	5.32	0.60	4.78	2.23
Female	6.65	3.73	6.15	4.45
EMPLOYMENT				
Race				
White	6.43	2.08		
Black	19.40	14.35		
Region				
Urban	14.40	11.89		
Rural	15.90	14.31		
Gender				
Male	9.89	9.08		
Female	16.75	17.08		

Notes: * Not statistically significant ($p > 0.05$) **R-squared per sub-group model is presented in brackets after the factor contribution of household size.

Source: NIDS wave 1 and wave 5; author's calculations. Post-stratified weights applied.

The contribution of employment to income inequality is greater in Black than White households. In 2017, employment status contributed 2.08% to income inequality in White HHs but 14.35% in Black households. A significant difference in the importance of employment status was also evident between male and female subgroups. The number of employed household members contributed 17.08% to income inequality in female-headed HHs compared to 9.08% in male-headed HHs in 2017. However, as discussed above, the contribution of employment status to income inequality is still lower than expected for all subgroups considering that earnings depend on employment. Race as a determining factor of income inequality has decreased over time for all subgroups. For example, for the female subgroup (Model 1), race contributed 24.13% in 2008 and 8.92% in 2017. This fall was also evident for the results for the South African population (Table 5.5).

Despite being statistically significant, the results for the subgroups need to be interpreted with caution given that the models used have different residual estimates with R-squared ranging from 0.2742 to 0.6086.

5.6 Conclusion

South African literature on the relationship between HC and socioeconomic inequalities most commonly uses an education-based measure to proxy HC. The results of this study suggest that this measurement approach is justified because in South Africa education attainment captures most of the effects of HC indicators (in comparison to schooling quality and health indicators) on income inequality. However, the study finds that the number of HH members employed in the labour market as well as the size of the HH are also important determinants of income inequality. HH size, although having moderate associations with income in the income generating model (between a 10% and 20% reduction in HH income per capita per additional person in the household), plays a large role in determining income inequality. The importance of more employed HH members is not as high as expected, contributing approximately 12% to HH income per capita inequality across both years. This result could be because of the number of HH members employed in low-paying jobs. Thus, rather than employment determining income inequality, higher levels of education attainment (and skills), subject to increasing income returns to education, are the dominant determining factor. Furthermore, social transfers (which are included in the HH income per capita estimate) could also reduce the impact of unemployment on income inequality.

Race continues to be a determining factor of income inequality in South Africa, but its importance is diminishing. Although, having a family member who has attended a quintile 5 school is not a clear determinant of income inequality, it does impact income inequality through school throughput rates and final educational attainment. A HH in which all members have good health is furthermore not directly found to contribute significantly to income inequality, but as with schooling quality the impacts could be indirect with good health contributing to greater participation in the labour force or to higher levels of education attainment.

It is thus concluded that improving school quality and higher levels of educational attainment are the surest means of reducing income inequality in South Africa for all races, for residents of different geographical areas, and for genders. Improving school attendance without improved schooling quality and higher levels of education attainment will not generate the desired results from education in reducing income inequality. This is a problem throughout Africa with the learning-adjusted years[100] of schooling in Sub-Saharan countries in the bottom half of a cross-country comparative study and more African countries performing below their predicted income level (Evans and Acosta, 2021). Without improvements in schooling quality, income inequality is expected to persist in South Africa or even to grow through disparities in the labour market. The findings are thus consistent with the conclusions drawn by van der Berg (2014) who highlighted that while, since the achievement of democracy, quality education services are benefiting a few wealthy members of society, the poor remain disadvantaged through poor service delivery. With higher educational attainment for all, HC measured using methods other than the education-based approach (such as education attainment proxies) may become more important. However, currently, using educational attainment as a proxy for

[100] A measure combining schooling quantity and quality (Filmer *et al.*, 2020)

HC is justified because education is the dominant HC factor determining income inequality. The results furthermore confirm that HC and social reproduction theory hold in South Africa. Investments in education will increase earnings, but the current education system is facilitating the reproduction of socioeconomic inequalities.

CHAPTER 6

6 CONCLUSION

The words of late president Nelson Mandela that "Education is the most powerful weapon which you can use to change the world" remains relevant in all countries. To achieve the goal of reducing poverty and inequality in South Africa by 2030, as envisaged in the "National Development Plan (NDP): Our future make it work", it is critical that the quality of education be improved and that human capital (HC) and capabilities be developed. Given the bidirectional relationship between HC and income inequality, HC development is required to make progress in reducing inequality and vice versa. Furthermore, the NDP includes the objective of inclusive development, which requires equity with regard to gender, race and spaces. These goals and objectives are all aligned with the Sustainable Development Goals set out by the United Nations.

To monitor the achievement of the NDP objectives it is however important for progress to be measured. This study focused on measuring HC and given that HC is an unobserved construct, measuring it presents challenges. The main research goal was to measure HC at a disaggregated level and to determine the importance of HC in explaining the high levels of inequality in South Africa. In order to achieve this goal, three subgoals were formulated:

Subgoal 1: To construct and compare a set of measures of HC at a disaggregated level in the South African context;
Subgoal 2: To measure the distribution of HC according to socioeconomic status in South Africa;
Subgoal 3: To determine the extent to which HC explains income inequality in South Africa.

6.1 The approach to achieving the goals

The research goals were achieved in four research articles, here presented in Chapters 2, 3, 4 and 5. Three methodological approaches were selected to achieve the goal of measuring HC at a disaggregated level: an education-based approach, a latent-variable approach, and by using a socioeconomic status human capital index (SES-HCI).

The education-based approach was included as a benchmark measure of HC (Chapters 3 and 5), given that education is widely considered to be a fundamental dimension of HC. Also, educational attainment has formed the basis of most HC research in South Africa because it is easily measured through available data (Lam, 1999; Liebbrandt *et al.,* 2010; Van der Berg, 2002, 2010; Pellicer and Ranchhod, 2012; Branson and Leibbrandt, 2013 and Wanka, 2014). The latent-variable approach was identified through a literature review study (Chapter 2) as an appropriate means to capture the unobserved nature of HC. The SES-HCI was employed to identify HC disaggregated by socioeconomic status and thus highlight within-country gaps in HC development (Chapter 4).

To estimate the distribution of HC, these different methods were applied to the 2017 Wave 5 NIDS data to estimate household/individual HC for the South African population and the distribution by socioeconomic indicators (with a focus on income, gender, race and geographical area). Furthermore, methodological challenges of measuring HC using each method were identified to assist future researchers who wish to analyse and monitor HC development and inequalities. Lastly, household income inequality in South Africa was decomposed according to the determinants of HC, using Fields' regression-based decomposition method (Chapter 5). The findings of each article are presented next, and the limitations of the research as well as recommended future research are discussed. Concluding remarks complete the work.

6.2 Summary of findings

To start, Chapter 2 focused on reconciling the terms 'meaning' and 'measurement' with regard to HC, in an effort to establish the meaning of HC and therefore how it should be measured. Initial exploration of the literature highlighted the potential for further development of measuring HC and the importance of measuring HC at a disaggregated level (individual, household or by subgroup) for the purpose of understanding the distribution of HC within countries.

Challenges in measuring HC were identified, taking into account the fact that both the definition and indicators to be included in the measure of HC are not definitive. HC was identified as statistically a latent variable which should therefore be measured as such. The earnings function, developed from HC theory and introduced originally in the 1950s by the microeconomic theorists Becker, Mincer and Schultz at the University of Chicago, was critiqued because it measures HC as the investments in HC that influence future real income rather than as an unobserved construct. The definition of HC as per the OECD (2001:18) "the knowledge, skills, competencies and attributes embodied in individuals that facilitate the creation of personal, social and economic well-being" was noted as encapsulating the meaning of HC. Sen's human capabilities approach was furthermore identified as a theoretical framework for supplementing the estimation and analysis of HC by broadening the vision of wellbeing. Data and methodological limitations were identified as problematic when measuring HC. Because of these difficulties, researchers measuring inequalities in South Africa have tended to estimate HC using an education-based approach, thus reducing the meaning of HC to education, and educational attainment in particular.

Recent literature, however, proposes a latent-variable approach to measuring HC, which has not previously been applied in South Africa. This emergent literature uses a structural equation modelling (SEM) framework and concludes that HC should be measured in a formative-reflective model. Its estimation models extend from factor analysis to path analysis, reduced rank regression, generalised redundancy analysis and a new least squared approach to extracting latent variables in a formative-reflective model. Furthermore, an actuarial method to estimate HC in monetary values has been developed for several countries. A thorough review of the different methods and their respective strengths and weaknesses was presented in Chapter 2. It was found that the variables to be considered in such a model should broadly fall under the headings of education, work experience, health, indicators linked to career progression and discrimination, and finally indicators reflective of HC, namely income. Given the intergenerational transfer of welfare, sociodemographic indicators should furthermore include parental socioeconomic status.

Previous studies concluded that a formative-reflective path model for the extraction of HC was necessary to ensure alignment with the definition of HC. It was however found that, given the statistical complexity and a lack of application of the methods beyond the Italian and the USA contexts, it would be challenging for future researchers to replicate the studies using household survey data. Furthermore, indicators included in a measure of HC should be refined according to data availability and specific contexts. The conclusion was that HC should be measured as a latent variable in a formative-reflective model. Therefore Chapter 3 aimed to apply the suggested approach using the South African NIDS dataset and educational attainment was used as a benchmark proxy measure of HC. The NIDS data was used throughout the remaining chapters, to allow for comparable results.

Following the conclusions synthesised in Chapter 2, a formative-reflective model within the framework of SEM was identified as theoretically appropriate for extracting a latent measure of human capital from household survey data. However, practical challenges highlighted in Chapter 2 became evident when applying the formative-reflective model using path analysis in South Africa. The first limitation was that certain variables could not be included, through a lack of observations, particularly for estimating on-the-job training and work experience in South African data. Secondly, when specified with South African data, the formative-reflective model did not converge to provide statistically significant results and was thus not applicable, since no clear path was identifiable.

A reflective confirmatory factor analysis model was however identified as appropriate to measure HC as a latent construct in South Africa. The reflective model was motivated since the indicators all represent the underlying construct, hence potential proxies of HC, and are correlated. The validity of the model was further proved by the empirical evidence of a strong bidirectional relationship between HC indicators and socioeconomic status in South Africa (Spaull, 2015). The theoretical frameworks in which the analysis of the latent construct HC in South Africa was conducted included microeconomic HC theory and social reproduction theory. A reflective model was thus specified, and indicators were selected considering the theoretical frameworks and empirical evidence of a persistent intergenerational transfer of poverty and inequality.

The latent HC variable was estimated, using the reflective model, and good model fit was identified. The distribution of the estimated HC latent variable (HC_L) provided insightful results and allowed for the comparison between the distribution of HC_L and HC_E (using educational attainment to proxy HC). Parental education attainment was found to be the highest factor influence on HC_L. When presenting the frequency distribution of HC_L in comparison to HC_E, a greater concentration of observations in lower levels of HC was evident when using the HC_L measure. Thus, the differences in the measures of HC_L and HC_E were driven, in large part, by the inclusion of parental education, schooling quality and education attainment as proxies associated with HC_L.

To profile HC by subgroup, the HC_L (using quintiles) and HC_E (using education attainment categories) were estimated by gender, geographical area, income quintile and race. A large concentration of HC_L in the bottom HC_L quintile was evident for income quintile 1 African rural dwellers. Male White income quintile 5 urban dwellers were found to have a higher concentration of HC_L in the top quintile. The HC_E education attainment categories supported the findings for HC_L, except when considering gender, where HC_E highlighted females as having higher levels of HC. However, disparities between race, geographical area and gender were not as pronounced using HC_E attainment categories as they were when comparing HC_L quintiles. It was thus noted that the disparities inherited from apartheid policies are likely to remain unless education quality and health service delivery improve. A limitation of the study was that the HC_E and HC_L measures by subgroup could not be compared using quintile estimates for both. This is because education attainment used in HC_E categories is not sufficiently precise to differ across population quintiles. The sensitivity analysis (in Chapter 3) achieved through path model modifications and an alternative estimation method, namely principal component analysis (PCA), identified robust results for the constructed measure of HC_L. Thus, estimating HC as a latent variable, in a SEM framework, is a viable alternative approach to measuring HC when analysing HC and the distribution thereof in South Africa.

Recommendations for future research include exploring additional path analysis models in the SEM framework within the South African context, particularly to identify an applicable formative-reflective model. However, it is expected that the problem of data convergence to estimate a latent variable, using a formative-reflective model, in assessing HC will persist until more precise household data measurements become available. It is also recommended that the earnings returns to HC_L be explored, given the higher earnings returns for higher levels of education in South Africa (Branson and Leibbrandt, 2012 and van Broekhuizen, 2011). A further proposal is for research focused on a better understanding of health as an indicator of HC in South Africa, and on supplementing the study of HC by considering the human capabilities approach, which highlights the importance of wellbeing beyond economic outcomes.

Chapter 4 took another approach to measuring HC, referring to the newly introduced HCI by the World Bank, which provides a method for capturing data on education, survival and health. Given the 2018 HCI score of 0.41 for South Africa and high levels of inequality, the importance of measuring productivity gaps in South Africa was highlighted. An interpretation of the HCI suggests that by the age of 18 South Africans born today can expect to earn on average only 41% of their full health and educational potential. It was noted that in 2018 South Africa ranked 126th out of the 157 countries for which an index score was estimated.

Data availability was identified as a limiting factor in estimating HCI by socioeconomic indicators in South Africa, particularly for the health and survival components. Using the NIDS dataset, the limited number of observations available for estimating adult survival and under-5 mortality rates meant that these indicators could not be included with an acceptable level of accuracy. The results from the 2017 NIDS dataset showed large distributional gaps between races, income quintiles and geographical areas, particularly when a graded curve for harmonised test scores (HTSs) was applied. The large impact of learning-adjusted school years on SES-HCI differences emphasises how poorly South Africa performs in terms of schooling quality.

A limitation of the HCI measurement identified was that it restricts estimates to the HC obtained by the age of 18. An extension of the SES-HCI was thus explored which estimated the schooling component using education attainment at age 25 years. When using a graded curve for HTS, the difference between the HCI estimates for income quintile 1 in comparison to quintile 5 (estimating education attainment at age 25 years old), and White compared to African individuals, was 19 percentage points. The difference between urban and rural residential areas was 8 percentage points (graded curve applied). These results suggest that to empower South Africans and reduce the gaps in HC, improved learning and skills, rather than time enrolled at school, should be emphasised. It was furthermore found that despite the SES-HCI estimates showing that women have higher HC than men, women have poorer employment prospects and earn less than men. Characteristics other than HC, such as additional penalties due to motherhood, were thus suggested as negatively impacting the labour market outcomes of women.

The research identified methodological considerations for future researchers and suggestions for methodological development. These include incorporating HC differences between working age individuals, thus expanding the definition of HC employed in the HCI. Furthermore, higher and increasing earnings returns to higher levels of education and applying returns to education and health, specific to different subgroups in South Africa, should be explored. Improved data availability, which allows for the data to be disaggregated further by using socioeconomic indicators, would allow for further analysis and monitoring of the distribution of HC. Despite data limitations, estimating the SES-HCI for various indicators provided important insights into the extent and causes of HC inequalities in South Africa. The SES-HCI estimates furthermore facilitate future research to monitor changes in HC gaps in South Africa.

Finally, Chapter 5 presented a research article which aimed to identify the extent to which HC contributes to income inequality. Fields' regression-based decomposition method was applied to the Wave 1 (2008) and Wave 5 (2017) NIDS datasets to identify changes over the 9-year period. The determinants of HC included were education, schooling quality and self-reported health. The analysis was conducted at the household level and a proxy for the educational attainment of the household was explored. It was revealed that using a different household reference person to measure household educational attainment affects the level of HC proxied by means of education attainment. Education attainment was noted as the most significant HC determinant contributing to income inequality which ranged from 30% to 42% of the predicted log of income per capita (depending on the model estimated).

The results thus suggest that education attainment as a proxy for HC is justified in South Africa because it captures, at the household level, most of the effects of HC indicators (in comparison to schooling quality and health indicators) on income inequality in the local context. With regards to schooling quality, although having a family member who has attended a quintile 5 school is not a clear determinant of income inequality, it does impact income inequality through school throughput rates and final educational attainment. A household in which all members have good health is furthermore not found directly to contribute significantly to income inequality, but in a similar manner to schooling quality, the impacts may be indirect, with good health contributing to greater participation in the labour force or to higher levels of educational attainment.

Household size and race were identified as important determinants of changes in income inequality between 2008 and 2017. Household size is noted as having an increasing impact, whereas the contribution of race decreased in those nine years.

Without future improvements in schooling quality, income inequality is expected to persist in South Africa, or even to grow, through disparities in the labour market. The findings are thus consistent with the conclusions drawn by van der Berg (2014) who highlighted that while, since the achievement of democracy, quality education services are benefiting a few wealthy members of society, the poor remain disadvantaged through poor service delivery. With higher educational attainment for all, HC measured using methods other than the education-based approach (such as education attainment proxies) may become more important.

The findings from all the studies demonstrate the importance of HC in understanding socioeconomic inequalities. Each of the measurement approaches to measuring HC provided valuable insights, as well as challenges. When decomposing the relationships between HC and income inequality, use of the education attainment approach was justified. However, valuable insights are gained when applying the SES-HCI and HC_L estimation of HC, particularly given the greater inequalities demonstrated in comparison to an education-based approach. The differences are highlighted when socioeconomic status and schooling quality are included in the measure of HC.

A common theme from the findings of all the studies was the need to improve education quality in South Africa. It is evident that without improvements in educational attainment and schooling quality an intergenerational transmission of inequality will continue and specific socioeconomic subgroups will continue to be disadvantaged and income inequality will persist.

6.3 Concluding remarks

Selecting the topic and research gap was a process of exploring a matter of universal and personal importance. One's education and personal abilities are one's greatest assets, which no person can take away. Without good health and healthcare, the value of other assets is reduced, even to the extreme of being meaningless. The socioeconomic environment in which we grow up is another major determinant of our life outcomes, particularly in South Africa.

This research provided the opportunity to highlight the multidimensional value embodied in each individual and its contribution to personal welfare. However, when education is recognised as the only proxy and economic wellbeing is the only recognised outcome measurement, challenges result in HC being ignored in wealth estimates. An initial consideration of the literature prompted the view that HC should be measured in a SEM framework, using a formative-reflective model. However, it was found that the lack of application of the approach and the available data allowed only for a limited estimate of HC using the SEM framework. The results from the study using SEM nonetheless provided insights into the distribution of HC, and the impact of the intergenerational transmission of inequalities. Breaking this cycle is imperative for future development.

Using the SES-HCI approach allowed for an approach which included consideration of additional health measures, including stunting and under-5 child mortality in HC estimates. It showed how HC is distributed in South Africa using the SES-HCI approach and the productivity gaps by socioeconomic indicators. The problem of poor-quality education and disparities in learning in South Africa was furthermore highlighted. What is HC, and how to measure it were again questioned and the importance of alternative approaches, allowing for the different dimensions of HC to be estimated, were highlighted.

Lastly, considering a household measure of HC and its impact on income inequality revealed how important higher levels of education attainment drive income inequality in South Africa, and showed that these higher levels of educational attainment are largely a result of the quality of education received. What is HC on a household level was questioned and education attainment was highlighted as capturing the impact of HC on income inequality when using a regression-based approach and the NIDS dataset.

It is clear that the stark economic inequalities in South Africa undermine democracy and the future of the country. The Corona virus pandemic, which has affected the systems in which we live in ways we could never imagine, has reemphasised this issue. This research reinforces the importance of good quality education for all citizens, regardless of race, gender, geographical location and socioeconomic background. Without this, the future of the country is bleak. With it, the future is bright and the possibilities limitless.

7 REFERENCES

ALAVI, M., VISENTIN, D. C., THAPA, D. K., HUNT, G. E., WATSON, R., and CLEARY, M. (2020). Chi-squared for model fit in confirmatory factor analysis. Journal of Advanced Nursing, 76: 2202-2211.

ARDINGTON, C. and GASEALAHWE, B. (2012). Health: Analysis of the NIDS Wave 1 and 2 Datasets. SALDRU Working paper number 80/NIDS discussion paper 2012/3. Cape Town: SALDRU.

ARENDSE, J.A. (2018). An investigation into the introduction of a new wealth tax in South Africa. PhD thesis. Makhanda: Department of Accounting, Faculty of Commerce, Rhodes University.

ARRIGHI, G., ASCHOFF, N, and SCULLY, B. (2008). Labor supplies in comparative perspective. The Southern African paradigm. Department of Sociology Working paper presented February 28. Berkley: Department of Sociology, University of California.

ARROW, K. J., DASGUPTA, P., GOULDER, L.H., MUMFORD K.J. and OLESON, K. (2012). Sustainability and the Measurement of Wealth. Environment and Development Economics, 17(03): 315-353.

ARROW, K. J., DASGUPTA, P., GOULDER, L.H., MUMFORD K.J. and OLESON, K. (2013). Sustainability and the measurement of wealth: further reflections. Environment and Development Economics, 18(4): 504– 516.

ASONGU, S.A. and NWACHUKWU, J.C. (2018). Educational quality thresholds in the diffusion of knowledge with mobile phones for inclusive human development in sub-Saharan Africa. Technological forecasting and social change, 129: 164-172.

AYYAK, M. and SEK, S.K. (2020). Decomposing Inequality in Household Consumption Expenditure in Malaysia. Economies, 8(83): 1-15.

BABYAK, M. A. and GREEN, S. B. (2010). Confirmatory factor analysis: An introduction for psychosomatic medicine researchers. Psychosomatic Medicine, 72(6): 587-597. Available at: https:/doi.org/10.1097/PSY.0b013e3181de3f8a [Accessed 10 January 2020]

BAGOZZI, R. (2011). Measurement and meaning in Information Systems and Organisational research: Methodological and Philosophical foundations. MIS Quarterly, 35(2): 261-292.

BARRO, R.J. and LEE, J.W. (1993). International comparisons of educational attainment. Journal of Monetary Economics, 32(3): 363–394.

BARRO, R.J. and LEE, J.W. (1996). International measures of schooling years and schooling quality. American Economic Review, 86(2): 218–223.

BASSIER, I. and WOOLARD, I. (2018). Exclusive growth: Rapidly increasing top incomes amidst low national growth in South Africa. REDI3x3 Working Paper 47. Cape Town: Research Project on Employment, Income Distribution and Inclusive Growth.

BECKER, G.S. (1962). Investing in Human Beings: A Theoretical Analysis. The Journal of Political Economy, 70(5): 9-49.

BECKER, G.S. (1964). Human Capital: A Theoretical and Empirical Analysis. New York: Columbia University Press for the National Bureau of Economic Analysis.

BLINDER, A.S. (1973). Wage discrimination: Reduced form and structural estimates. Journal of Human Resources, 8: 436–55.

BLUME, L.E. and DURLAUF, S.N. (2015). Capital in the Twenty-First Century: A Review Essay. Journal of Political Economy, 123(4): 749-777.

BOUND, J. (1991). Self-reported versus objective measures of health in retirement models. Journal of Human Resources, 26(1): 106-138.

BOURDIEU, P. and PASSERON, J.C. (1977). Reproduction in education, society and culture. Beverly Hills: Sage.

BOWLES, S. and GINTIS, H. (1975). The Problem with Human Capital Theory--A Marxian Critique. The American Economic Review, 65(2): 74-82.

BRANSON, N., ARDINGTON, C., LAM, D. and M. LEIBBRANDT, M. (2013). Changes in education, employment and earnings in South Africa – A cohort analysis. SALDRU Working Paper 105. Cape Town: Southern Africa Labour and Development Research Unit.

BRANSON, N. and LEIBBRANDT, M. (2013). Educational Attainment and Labour Market Outcomes in South Africa, 1994-2010. OECD Economics Department Working Papers, No. 1022. Paris: OECD Publishing. Available at: http://doi.org/10.1787/5k4c0vvbvv0q-en [Accessed 10 January 2020]

BREEN, R., and JOHNSON, J. (2005). Inequality of opportunity in comparative perspective: recent research on educational attainment and social mobility. Annual Review of Sociology, 31(1): 223–243. Available at:

https://doi.org/10.1146/annurev.soc.31.041304.122232 [Accessed 10 January 2019]

BROPHY, T., BRANSON, N., MLATSHENI, C. and WOOLARD, I. (2018). National Income Dynamics Study Panel User Manual. Cape Town: National Income Dynamics Study, University of Cape Town. Available at: http://www.nids.uct.ac.za/images/documents/20180831-NIDS-W5PanelUserManual-V1.0.pdf [Accessed 10 January 2019]

BUDLENDER, J., LEIBBRANDT, M. and WOOLARD, I. (2015). South African poverty lines: A review and two new money-metric thresholds. SALDRU Working Paper 151. Cape Town: Southern Africa Labour and Development Research Unit.

CASALE, D. and POSEL, D. (2005). Women and the Economy: How Far Have We Come? Agenda: Empowering Women for Gender Equity, 64: 21-29.

CASTELLÓ, A. and DOMENECH, R. (2002). Human Capital, Inequality and Economic Growth: Some New Evidence. The Economic Journal, 112(March): C187-C200.

CASTELLÓ-CLIMENT, A. and DOMENECH, R. (2017). Human Capital and Income Inequality: New Fact and Explanations. University of Valencia. Available at: https://www.uv.es/rdomenec/Human_Income_Inequality_6May2017.pdf [Accessed 1 February 2019]

CHANI, M.I., JAN, S.A., PERVAIZ, Z. and CHAUDHARY, A.R. (2014). Human capital inequality and income inequality: testing for causality. Quality and Quantity, 48: 49-156.

CHANT, S. (2007). Gender, Generation and Poverty. Cheltenham: Edward Elgar.

CHATTERJEE, A., CZAJKA, L. and GETHIN, A. (2020). Estimating the distribution of household wealth in South Africa. WID.world working paper no 2020/06. Available at: https://www.wider.unu.edu/publication/estimating-distribution-household-wealth-south-africa [Accessed 4 March 2021].

CHECCHI, D. (2001). Education, Inequality and income inequality. Distributional Analysis Research Programme (DARP) Discussion Paper No 52. London: London School of Economics.

CHIAPPA, R. and MEJIAS R.R. (2019). Unfolding the direct and indirect effects of social class of origin on faculty income. Higher Education, 78: 529–555. Available at: https://doi.org/10.1007/s10734-019-0356-4 [Accessed 10 January 2019]

CHINHEMA, M., BROPHY, T., BROWN, M., LEIBBRANDT, M., MLATSHENI, C. and WOOLARD, I. (eds) (2016). NIDS Panel User Manual 2016. Cape Town: NIDS, University of Cape Town. Available at: http://www.nids.uct.ac.za/images/documents/wave4/20170227-NIDS-W4PanelUserManual-V1.1.pdf [Accessed 1 March 2019]

CHIU, W.H (1998). Income inequality, human capital accumulation and economic performance. The Economic Journal, 108(January): 44-59.

CIPD (2017). Human capital theory: Assessing the evidence for the value and importance of people to organizational success. Technical Report May 2017. London: CIPD.

COLTMAN, T., DEVINNEY, T. M., MIDGLEY, D. F. and VENAIK, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. Journal of Business Research, 61(12): 1250-1262. Available at: http://ro.uow.edu.au/infopapers/689 [Accessed 3 March 2021].

COSTA, M. (2006). The Dagum Model of Human Capital Distribution. STATISTICA, LXVI(3): 313-324

DAGUM, C. (1994). Human capital, income and wealth distribution models and their applications to the USA. Proceedings of the 154th Meeting of the American Statistical Association, 253–258. Virginia, USA: American Statistical Association.

DAGUM, C. and SLOTTJE, D.J. (2000). A new method to estimate the level and distribution of the household human capital with application. Structural Change and Economic Dynamics, 11: 67–94.

DAGUM, C. and VITTADINI, G. (1996). Human capital measurement and distributions. Proceedings of the Business and Economic Statistics Section, American Statistical Association, 194–199. Virginia, USA: American Statistical Association.

DAGUM C., VITTADINI, G., LOVAGLIO, P.G. and COSTA, M. (2003). A microeconomic recursive model of human capital, income and wealth determination: Specification and estimation. Joint Statistical Meetings – Business and Economics Statistics Section, 2003: 1119-1126.

DAGUM C., VITTADINI, G. and LOVAGLIO, P.G. (2007). Formative indicators and effects of a causal model for household human capital with applications. Econometric Review, 26(5): 579–596.

DE GREGORIO, J., and LEE, J.W. (2002). Education and income inequality: New evidence from cross-country data. The Review of Income and Wealth, 48(3): 395-416.

DENG, Q. and LI, S. (2009). What lies behind rising earnings inequality in urban China? Regression-based decompositions. CEDSifo Economic Studies, 55(3-4): 598-623.

DEPKEN, C., CHISENI, C. and ITA, E. (2019). Returns to Education in South Africa: Evidence from the National Income Dynamic Study. International Review of Economics and Business, 22(1): 1-12.

DESALVO, K., BLOSER, N., REYNOLDS, K., HE, J. and MUNTNER, P. (2005). Mortality Prediction with a Single General Self-Rated Health Question: A Meta-Analysis. Journal of General Internal Medicine, 20: 267–275.

DI BARTOLO, A. (1999). Human Capital estimation through Structural Equation Modelling with some categorical observed variables. International Workshop on Correlated Data: estimating function approach, October. Italy: University of Milano-Bicocca.

D'SOUZA, R., GATTI, R. and KRAAY, A. (2019). A socioeconomic disaggregation of the World Bank Human Capital Index. Policy Research Working Paper 2020. Washington, D.C., United States of America: World Bank Group.

DU TOIT, C. (2008). Rethinking South Africa's macro-economic policy: Integrated social development as a key to accelerated and shared growth. Presentation to the Human Sciences Research Council/Development Bank of Southern Africa Policy dialogue conference: The potentials for and challenges of constructing a democratic state in South Africa, Magaliesberg, 4-6 June.

EISNER, R. (1985). The total incomes system of accounts. Survey of Current Business, 65(1): 24–48.

EVANS, P.B. (2010). Constructing the 21st Century Developmental State: Potentialities and Pitfalls. Cape Town: HSRC Press. Available at: www.hsrcpress.ac.za [Accessed 23 January 2021]

EVANS, D. and ACOSTA, A.M. (2021). Education in Africa: What are we learning? Journal of African Economies, 30(1): 13–54.

FARR, W. (1853). Equitable taxation of property. Journal of the Royal Statistical Society, XVI(March): 1–45.

FATTORE, M., PELAGATTI, M. and VITTADINI. G. (2018). A least squares approach to latent variables extraction in formative–reflective models. Computational Statistics and Data Analysis, 120(2018): 84–97.

FEDDERKE, J., DE KADT, R. and LUIZ, J. (2003). Capstone or deadweight? Inefficiency, duplication and inequity in South Africa's tertiary education system, 1910—93. Cambridge Journal of Economics, 27(3): 377- 400.

FERNANDEZ, R. and ROGERSON, R. (1992). Human Capital Accumulation and Income distribution. NBER Working Papers Series No. 3994. Cambridge, MA, USA: NBER.

FIELDS, G.S. (2003). Accounting for income inequality and its change: A new method, with application to the distribution of earnings in the United States. Research in Labor Economics, 22: 1-38.

FINNOFF, K. (2015). Decomposing inequality and poverty in post-war Rwanda: The roles of gender, education, wealth and location. Development Southern Africa, 32(2): 209-228.

FIORIO, C.V. and JENKINS, S.P. (2007). "ineqrbd": Regression-based inequality decomposition, following Fields (2003) UKSUG. Available at: https://www.stata.com/meeting/13uk/fiorio_ineqrbd_UKSUG07.pdf [Accessed 1 June 2020]

FOLDVARI, P., and VAN LEEUWEN, B. (2005). An estimation of the Human Capital stock in Eastern and Central Europe. Eastern European Economics, 43(6): 53–65.

FOLLONI, G. and VITTADINI, G. (2010). Human Capital Measurement: A Survey. Journal of Economic Surveys, 24(2): 248-279.

FRANSMAN, T. and YU, D. (2018). Multidimensional poverty in South Africa in 2001-16. Development Southern Africa, 29(1): 19-34.

GROSSMAN, M. (2000). The Human Capital Model. Handbook of Health Economics, 1: 347-408.

GU, W. and MACDONALD. R. (2016). Human capital: Satellite Account. Statistics Canada. Available: https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.20/2016/Item_7_Canada_UNECE_Human_Capital.pdf [Accessed 10 October 2020]

HALL, K., RICHTER, L., MOKOMANE, Z. and LAKE, L. (2018). South African Child Gauge 2018. Cape Town: Children's Institute, University of Cape Town.

HODGSON, G.M. (2015). What is capital? Economists and sociologists have changed its meaning: should it be changed back? Cambridge Journal of Economics, 38(5): 1063-1086.

HOFMEYER, H. (2018). Home background and schooling outcomes in South Africa: Insights from the National Income Dynamics Study. Stellenbosch Economics Working Papers: WP01/2018. Stellenbosch: The Department of Economics and the Bureau for Economic Research, University of Stellenbosch.

HOWIE, S.J., COMBRINK, C., ROUX, K., TSHELE, M., MOKOENA, G.M. and PALANE, N.M, (2017). PIRLS Literacy 2016 Progress in International Reading Literacy Study 2016: South African Children's Reading Literacy Achievement. Pretoria: Centre for Evaluations and Assessment.

HUNDENBORN, J., LIEBBRANDT, L. and WOOLARD, I. (2016). Drivers of Inequality in South Africa, A Southern Africa Labour and Development Research Unit Working Paper 194. Cape Town: SALDRU, University of Cape Town.

JARVIS, C.B., MACKENZIE, S.B. and PODSAKOFF, P.M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. Journal of Consumer Research, 30: 199-218.

JOHANSEN, A.L. (2014). The effect of human capital on income inequality. Masters Thesis. Denmark: Department of Economics, Copenhagen Business School.

KAMANZI, A., MCKAY, A., NEWEL, A., RIENZO, C. and TAFESSE, W. (2021). Education, Access to Better Quality Work and Gender: Lessons from the Kagera Panel Data Set. Journal of African Economies, 30(1): 103–127. Available at: https://doi.org/10.1093/jae/ejaa011 [Accessed 16 August 2020]

KEETON, G. (2014). Inequality in South Africa. The Journal of the Helen Suzman Foundation, 74: 26-31.

KLENOW, P. and RODRIGUEZ-CLARE, A. (1997). The Neoclassical Revolution in Growth Economics: Has It Gone Too Far? NBER Macroeconomics Annual. Chicago: NBER, The University of Chicago Press.

KRAAY, A. (2018). Methodology for the World Bank Human Capital Index. Policy Research Working Paper 8593. Washington, D.C., USA: World Bank.

KREIDER, B. (1999). Latent work disability and reporting bias. Journal of Human Resources, 34(4): 734-769.

KRIPPNER, G. (2005). The Financialization of the American Economy. Socio-Economic Review, 3(2): 173-208.

KRIPPNER, G. (2011). Capitalizing on Crisis: The Political Origins of the Rise of Finance. Cambridge, MA: Harvard University Press.

LAM, D. (1999). Generating Extreme Inequality: Schooling, Earnings, and Intergeneration Transmission of Human Capital in South Africa, Population Studies Centre Research Report. Michigan: Michigan University, Population Studies Centre.

LAM, D., FINN, A. and LEIBBRANDT, M. (2015). Schooling Inequality, Returns to Schooling, and Earnings Inequality: Evidence from Brazil and South Africa. WIDER Working Paper 2015/050. Helsinki: UNU-WIDER.

LAVERDE-ROJAS H, CORREA J.C., JAFFE K. and CAICEDO M.I. (2019). Are average years of education losing predictive power for economic growth? An alternative measure through structural equations modelling. PLoS ONE, 14(3), e0213651. Available at: https://doi.org/10.1371/journal.pone.0213651 [Accessed 10 January 2019]

LE, T., GIBSON, J. and OXLEY, L. (2003). Cost and income based measures of human capital. Journal of Economic Surveys, 17(3): 271–305.

LE, T., GIBSON, J. and OXLEY, L. (2005). Measurements of Human Capital: A review of the literature. New Zealand Treasury Working Paper 05/10. Wellington, New Zealand: New Zealand Treasury.

LEE, J-W. and LEE, H. (2018). Human Capital and Income Inequality. Asian Development Bank Institute Working Paper 810, Asian Development Bank Institute. Tokyo: Asian Development Bank Institute

LEIBBRANDT, M., WOOLARD, I., MCEWEN, H. and KOEP, C. (2010). Employment and Inequality outcomes in South Africa. Southern Africa Labour and Development Research Unit and School of Economics. Cape Town: University of Cape Town.

LEIBBRANDT, M., FINN, A. and WOOLARD, I. (2012). Describing and decomposing post-apartheid income inequality in South Africa. Development Southern Africa, 29(1): 19-34.

LIM, S.S., UPDIKE, R.L., KALDJIAN, A.S., BARBER, R.M., COWLING, K., YORK, H., FRIEDMAN, J., XU, R., WHISNANT, J.L., TAYLOR, H.J., LEEVER, A.T., ROMAN, Y., BRYANT, M.F., DIELEMAN, J., GAKIDOU, E., MURRAY, C.J.L. (2018).

Measuring human capital: a systematic analysis of 195 countries and territories, 1990–2016. The Lancet, 392: 1217 - 1234. doi: 10.1016/S0140-6736(18)31941-X.

LIMANLI, O. (2017). Accounting for Income Inequality in Turkey: Regression Based Decomposition Approach. Journal of Academic Researchers and Studies, 9: 137–46.

LIU, G.AND FRAUMENI, B. (2021). Introduction. In Fraumeni, B. (ed). Human Capital Measurement. Cambridge: Academic Press, Elsevier. pp. ix - xix.

LOEHLIN, J. C. and BEAUJEAN, A. A. (2017). Latent Variable Models: An Introduction to Factor, Path, and Structural Equation Analysis. New York, NY: Routledge. Available at: https://doi.org/10.4324/9781315643199 [Accessed 10 January 2019]

LOVAGLIO, P.G. (2008). Process of accumulation of Italian human capital. Structural Change and Economic Dynamics, 19: 342-356.

LOVAGLIO, P.G. (2010). The estimation of human capital by administrative archives in a static and longitudinal perspective: The case of Milan. Journal of Economic Surveys, 24(2): 280–308.

LOVAGLIO, P.G. and FOLLONI, G. (2011). The estimation of Human Capital in structural models with flexible specification. Paper presented at International Conference on Human Capital and Employment in the European and Mediterranean Area, March 2011. Italy: Bologna.

LOVAGLIO, P.G., and VITTADINI, G. (2014). Structural Equation Models in a Redundancy Analysis Framework With Covariates. Multivariate Behavioral Research, 49(5): 486-501. Available at: https://doi.org/10.1080/00273171.2014.931798 [Accessed 24 March 2020]

LUNDBORG, P., NORDIN, M. and ROOTH, D.O. (2018). The intergenerational transmission of human capital: the role of skills and health. Journal of Population Economics, 31: 1035–1065. Available at: https://doi.org/10.1007/s00148-018-0702-3 [Accessed 10 January 2021]

LUXEMBOURG INCOME STUDY (LIS) DATABASE (2019). Luxembourg income study database. Available at: https://www.lisdatacenter.org. [Accessed 5 July 2020]

MAGADLA, S., LEIBBRANDT, M. and MLATSHENI, C. (2019). Does a Motherhood Penalty Exist in the Postapartheid South African Labour Market? Working Paper 247 Version 1 NIDS Discussion Paper 2019/14. Cape Town: SALDRU

MAHOON, H., RADENY, M., KINYANGI, J. and CRAMER, L. (2003). Climate change vulnerability and risk assessment of agriculture and food security in Ethiopia: Which way forward. CCAFS Working Paper no. 59. Denmark: CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS).

MANNA, R. and A. REGOLI (2012). Regression-based approaches for the decomposition of income inequality in Italy, 1998–2008. Rivista di statistica u_ciale, 14: 5–18.

MARANDU, S. (2011). Full life tables for South Africa from vital registration data, 2006-2008. Masters thesis. Cape Town: Centre of Actuarial Research, Faculty of Commerce, UCT.

MARSHALL, A. (1890). Principles of Economics. London: Macmillan (Reprinted 1920).

MARX, K. (1977). *Capital* Volume 1. New York: Vintage

MILLER, L.C., JOSHI, N., LOHANI, M., ROGERS, B., MAHATO, S., GHOSH, S. and WEBB, P. (2017). Women's education level amplifies the effects of a livelihoods-based intervention on household wealth, child diet, growth in rural Nepal. International Journal of Equity in Health, 16: 183. Available at: https://doi.org/10.1186/s12939-017-0681-0 [Accessed 17 April 2020]

MINCER, J. (1958). Investment in Human Capital and Personal Income Distribution. Journal of Political Economy, 66(4): 281-302.

MINCER, J. (1974). Schooling, Experience, and Earnings. New York: NBER Press.

MOOI E., SARSTEDT M. and MOOI-RECI I. (2018). Principal Component and Factor Analysis. In: Market Research. Springer Texts in Business and Economics. Singapore: Springer. Available at: https://doi.org/10.1007/978-981-10-5218-7_8 [Accessed 10 January 2020]

MORDUCH, J. and SICULAR, T. (2002). Rethinking Inequality Decomposition with Evidence from Rural China. Economic Journal, 112: 93–106.

MOSOMI, J. (2019). Distribution changes in the gender wage gap in post-apartheid South African labour market. WIDER Working paper 17/2019. Finland: UNU-WIDER.

NATIONAL PLANNING COMMISSION (2012). Our future make it happen - The National Development Plan 2030. Executive summary. Available at:

https://www.gov.za/issues/national-development-plan-2030 [Accessed 1 November 2020]

NDoH, STATSSA, SAMRC, and ICF (2019). South Africa Demographic and Health Survey 2016. Pretoria, South African and Rockville, Maryland, USA: NDoH, StatsSA, SAMRC, and ICF.

NIDS (2018). What is NIDS, National Income Dynamics Survey. Available at: http://www.nids.uct.ac.za/about/what-is-nids [Accessed 5 January 2019]

NIDS (2020). Data/About NIDS FAQs. National Income Dynamics Survey. Available at: http://www.nids.uct.ac.za/nids-data/documentation/faqs/data-about-nids [Accessed 1 June 2020]

NWOSU, C.O. (2015). An analysis of the relationship between health and the labour market in South Africa. PhD thesis. Cape Town: Faculty of Commerce, University of Cape Town.

OAXACA, R. (1973). Male–female wage differentials in urban labor markets. International Economic Review, 14: 693–709.

OLANIYAN, D.A. and T. OKEMAKINDE (2008). Human Capital Theory: Implications for Educational Development. European Journal of Scientific Research. 24 (2): 157-162.

ORGANISATION for ECONOMIC CO-OPERATION and DEVELOPMENT (OECD) (1998). Human capital investment: an international comparison. Paris: Centre for International Research and Innovation. Available at: https://www.oecd-ilibrary.org [Accessed 1 June 2020]

ORGANISATION for ECONOMIC CO-OPERATION and DEVELOPMENT (OECD) (2001). The wellbeing of nations: The role of human and social capital. Paris: Centre for Educational Research and Innovation. Available at: https://www.oecd-ilibrary.org [Accessed 1 June 2020]

ORGANISATION for ECONOMIC CO-OPERATION and DEVELOPMENT (OECD) (2014). Education at a Glance 2014: OECD Indicators. Paris: OECD Publishing. Available at: https://doi.org/10.1787/eag-2014-en [Accessed 10 January 2020]

ORTHOFER, A. (2017). Savings and Wealth in the Context of Extreme Inequality. PhD thesis. Stellenbosch: Faculty of Economics and Management Sciences, Stellenbosch University.

OTTERBACH, S. and ROGAN, M. (2017). Spatial differences in stunting and household agricultural production in South Africa: (re-)examining the links using national panel survey data. Hohenheim discussion paper 13-2017. Germany: University of Hohenheim.

OXLEY, L., LE, T, and GIBSON, J. (2008). Measuring human capital: alternative methods and international evidence. Korean Economic Review, 24(2): 283–344.

PATRINOS, H. and ANGRIST, N. (2018). Global dataset on Education Quality. World Bank Policy Research Working Paper 8592. Washington, D.C., USA: World Bank.

PANDEY, A. (2013). Inequality in consumption expenditure at household level in Uttar Pradesh: Regression Based Inequality Decomposition (RBID). Journal of Global Economy, 9: 135–152.

PELLICER, M. and RANCHHOD, V. (2012). Inequality Traps and Human Capital Accumulation in South Africa. A Southern Africa Labour and Development Research Unit Working Paper 86. Cape Town: SALDRU, University of Cape Town.

PETT, M., LACKEY, N., and SULLIVAN, J. (2003). Making sense of factor analysis. Thousand Oaks: Sage Publications, Inc.

PETTY, W. (1690). Political Arithmetik, or a Discourse Conceirning the Extent and Value of Lands, People, Buildings. Reprint in: Hull, C.J. 1899. The Economic Writings of Sir William Petty. Cambridge: Cambridge University Press.

PIETERS, J. (2011). Education and Household Inequality Change: A decomposition analysis for India. Journal of Development Studies, 47(12): 1909-1924.

PIKETTY, T. (2014). Capital in the Twenty-First Century. Cambridge: Harvard University Press.

POSEL, D., and VAN DER STOEP, G. (2014). Co-resident and absent mothers: Motherhood and labour force participation in South Africa. Available at: https://mpra.ub.uni-muenchen.de/id/eprint/52907 [Accessed 8 December 2020]

RANI, U. and FURRER, M. (2017). Decomposing income inequality into factor income components. ILO Research Paper no. 15. Geneva, Switzerland: International Labour Office.

REDDY, V., VISSER, M., WINNAAR, L., ARENDS, F., JUAN, A., PRINSLOO, C.H. and ISDALE, K. (2016). TIMSS 2015: Highlights of Mathematics and Science Achievement of Grade 9 South African Learners. Pretoria: HSRC.

ROGAN, M. (2016). Qualitative perceptions of the meaning of headship and female-headed households in post-apartheid South Africa. Social Dynamics, 42(1): 175-195. Available at: http10.1080/02533952.2016.1166550

SCHULTZ, T.W. (1959). Investment in man: an economist's view. Social Service Review, 33(2): 109–117.

SCHULTZ, T.W. (1961). Investment in human capital. American Economic Review, 51(1): 1–17.

SEN A. (1999). Development as a freedom. New York: Knopf.

SHAHPARI, G., and DAVOUDI, P. (2014). Studying effects of human capital on income inequality in Iran. Social and Behavioural Science, 109(2014): 1386 – 1389.

SHORROCKS, A. (1982). Inequality decomposition by factor components. Econometrica, 50(1): 193-211.

SLOTTJE, D. (2010). Human Capital Measurement: Theory and Practise. Journal of Economic Surveys, 24(2): 201-205.

SOUTHERN AFRICAN LABOUR and DEVELOPMENT RESEARCH UNIT (SALDRU) (2017). National Income Dynamics Study, Wave 5 [dataset]. Version 1.0.0 Pretoria: Department of Planning, Monitoring, and Evaluation [funding agency]. Cape Town: SALDRU [implementer], 2018. Cape Town: DataFirst [distributor], 2018. Available at: https://doi.org/10.25828/fw3h-v708 [Accessed 5 January 2020]

SPAULL, N. (2015). Schooling in South Africa: How low-quality education becomes a poverty trap. South African Child Gauge, 12: 34-41.

SPAULL, N. (2019). Equity: A Price too high to pay? In Spaull, N. and Jansen. D. (eds). South African Schooling: The Engine of Inequality. Switzerland: Springer Nature, pp. 1-24.

SPAULL, N. (2020). Putting golf club needs before social housing is one way the rich fail the poor. Business Live, 22 January 2020. Available at: https://www.businesslive.co.za/bd/opinion/2020-01-22-putting-golf-club-needs-before-social-housing-is-one-way-the-rich-fail-the-poor/ [Accessed 5 January 2021]

SPAULL, N. AND KOTZE, J. (2015). Starting behind and staying behind in South Africa. The case of insurmountable learning deficits in mathematics. International Journal of Educational Development, 41(2015): 13-24.

STATACORP (2013). Stata: Release 13. Statistical Software. College Station, Texas, United States of America: StataCorp LP.

STATSSA (2011). Census 2011: A profile of education enrolment, attainment and progression in South Africa. Report No. 03-01-812011. Pretoria: StatsSA.

STATSSA (2017). Poverty trends in South Africa: An examination of absolute poverty between 2006 and 2015. Report 03-10-06. Pretoria: Statistics South Africa.

STATSSA (2019). Inequality trends in South Africa, Report No. 03-10-19. Pretoria: Statistics South Africa.

STATSSA (2021). Quarterly labour force survey, Report No. P0211. Pretoria: Statistics South Africa. STIGLITZ, J.E., SEN, A.D. and FITOUSSI, J.-P. (2009). Report by the Commission on the Measurement of Economic Performance and Social Progress. Available at: http://www.stiglitz-senfitoussi.fr/documents/rapport_anglais.pdf. [Accessed 5 January 2021]

SUHRCKE, M., MCKEE, M., STUCKLER, D., SAUTO ARCE, R., TSOLOVA, S. and MORTENSEN J. (2006). The contribution of health to the economy in the European Union. Public Health. 120(11): 994-1001. Available at: https://doi.org/10.1016/j.puhe.2006.08.011. [Accessed 10 January 2021]

SUMMERS, L.H. (2014). The Inequality Puzzle. Democracy Journal, 32: 1-7.

TAO, H.L. and STINSON, T.F. (1997). An Alternative Measure of Human Capital Stock. University of Minnesota Economic Development Center Bulletin 97/01. Minnesota, Unites States: University of Minnesota.

TAYLOR, S. (2019). How can learning inequalities be reduced? Lessons learnt from experimental research in South Africa. In Spaull, N. and Jansen. D. (eds). South African Schooling: The Engine of Inequality. Switzerland: Springer Nature, pp. 321-336.

TEAL, F. (2021). Education and Labour Market Outcomes in sub-Saharan Africa. Journal of African Economies, 30(1): 1–12. Available at: https://doi.org/10.1093/jae/ejaa014 [Accessed 17 April 2020]

THE ECONOMIST (2019). South Africa’s youngsters are let down by lousy education system. The Economist, April 25th 2019. Available at: https://www.economist.com/special-report/2019/04/25/south-africas-youngsters-are-let-down-by-a-lousy-education-system [Accessed 8 December 2020.

THOMSON, K.C., GUHN, M., RICHARDSON, C.G. and SHOVELLER, J.A. (2017). Associations between household educational attainment and adolescent positive mental health in Canada. SSM-Population health, 3(2017): 403-410.

TIMAEUS, I.N., SIMELANE, S. and LETSOALO, T. (2013). Poverty, Race, and Children's Progress at School in South Africa. The Journal of Development Studies, 49(2): 270-284.

TREIBLMAIER, H. and FILZMOSER, P. (2010). Exploratory factor analysis revisited: How robust methods support the detection of hidden multivariate data structures in IS research. Information and Management, 47(4): 197-207.

TRIPATHI, S. (2017). Source of Inequality in Consumption Expenditure in India: A Regression-Based Inequality Decomposition Analysis. Journal of Income Distribution, 26: 1–34.

UNITED NATINS (2015). Transforming our world: The 2030 Agenda for Sustainable Development. Available at: https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf [Accessed 3 March 2021].

UNITED NATIONS (2016). Guide on Measuring Human Capital, Conference of European Statisticians. Norway: United Nations Economic Commission for Europe CE/CES/2016/2/Add.2.

UNITED NATIONS DEVELOPMENT PROGRAMME (1990). Human Development Report 1990: Concept and Measurement of Human Development. New York. Available at: http://www.hdr.undp.org/en/reports/global/hdr1990. [Accessed 30 June 2021]

UNITED NATIONS DEVELOPMENT PROGRAMME (2020). Human Development Index. UNDP. Available at: http://hdr.undp.org/en/content/human-development-index-hdi [Accessed 28 May 2020]

VAN BROEKHUIZEN, H. (2011). Labour Market Returns to Education Attainment, School Quality and Numeracy, in South Africa. Masters thesis. Stellenbosch: Faculty of Economics and Management Sciences, Stellenbosch University.

VAN DER BERG, S. and LOUW, M. (2004). Changing patterns of South African income distribution: Towards time series estimates of distribution and poverty. South African Journal of Economics, 72(3): 546–72.

VAN DER BERG, S. (2002). Education, Poverty and Inequality in South Africa, Paper presented at the conference of the Centre for the Study of African Economies on Economic growth and poverty in Africa, March 2002. Stellenbosch, South Africa: University of Stellenbosch.

VAN DER BERG, S. (2010). Current poverty and income distribution in the context of South African history. Stellenbosch Economic Working Papers: 22/10. Stellenbosch: Department of Economics, University of Stellenbosch.

VAN DER BERG, S. (2014). Inequality, poverty and prospects for redistribution. Development Southern Africa, 31(2): 197-218. Available at http://doi.org/10.1080/0376835X.2013.871196 [Accessed 5 January 2020]

VAN DER BERG., S. and GUSTAFSSON, M. (2019). Educational outcomes in post-apartheid South Africa: Signs of progress despite great inequality. In Spaull, N. and Jansen. D. (eds). South African Schooling: The Engine of Inequality. Switzerland: Springer Nature, pp. 25-46.

VICTORIA, C.G., BARROS, A.J., BLUMENBERG, C., COSTA, J.C., VIDALETTI, L.P., WEHRMEISTER, F.C., MASQUELIER, B., HUG, L. and YOU, D. (2020). Association between ethnicity and under-5 mortality rate. Lancet Glob Health, 2020(8): e352-61.

VITTADINI, G., DAGUM, C., LOVAGLIO, P.G. and COSTA, M. (2003). A method for the estimation of the distribution of human capital from sample surveys on income and wealth. 2003 Joint Statistical Meetings - Business & Economic Statistics Section, pp. 4381–4388.

VITTADINI, G. and LOVAGLIO, P.G. (2007). Evaluation of the Dagum-Slottje method to estimate Human Capital. Structural Change and Economic Dynamics, 18(2007): 270–278.

WANKA, F. (2014). The impact of educational attainment on household poverty in South Africa: A case study Limpopo province. Masters thesis. Cape Town: Department of Economics, University of Western Cape.

WEISBROD, B. A. (1961). The valuation of human capital. Journal of Political Economy, 69(5): 425–436.

WENDELSPIESS CHÁVEZ JUÁREZ, F. (2015). Measuring Inequality of Opportunity with Latent Variables. Journal of Human Development and Capabilities, 16(1): 106-121.

WILLIS, RJ. (1986). Wage determinants: a survey and reinterpretation of human capital earnings functions. In Ashenfelter, O., and Layard, R. Handbook of Labor Economics, Vol I. New York: Elsevier, pp. 525-602.

WößMANN L. (2003). Specifying human capital. Journal of Economic Surveys, 17(3): 239–270.

WORLD BANK (2012). South Africa Economic Update: Focus on Inequality of Opportunity. World Bank, Washington DC.

WORLD BANK (2018a). The human capital project. Available at: http://hdl.handle.net/10986/30498 [Accessed 5 January 2020]

WORLD BANK (2018b). South Africa. World Bank. Available at: https://databank.worldbank.org/data/download/hci/HCI_2pager_ZAF.pdf [Accessed 26 February 2019]

WORLD BANK (2019a). GINI index (World Bank estimate) – South Africa. World Bank. Available at: https://data.worldbank.org/indicator/SI.POV.GINI?locations=ZA [Accessed 5 January 2020]

WORLD BANK (2019b). GINI index (World Bank estimate) – Vietnam. World Bank. Available at: https://data.worldbank.org/indicator/SI.POV.GINI?locations=VN [Accessed 3 March 2020]

WORLD BANK (2019c). Full HCI Dataset. World Bank. Available at: https://www.worldbank.org/en/publication/human-capital#Data [Accessed 16 June 2020]

WORLD BANK (2019d). African Human Capital Plan. World Bank. Available at: https://www.worldbank.org/en/region/afr/publication/africa-human-capital-plan [Accessed 10 November 2020]

WORLD ECONOMIC FORUM (2017). The Global Human Capital Report 2017 - Preparing People for the Future of Work. Available at: _https://www.weforum.org [Accessed 30 June 2021]

ZIZZAMIA, R., SCHOTTE, S. and LEIBBRANDT, M. (2018). Snakes and Ladders and Loaded Dice: Poverty dynamics and inequality in South Africa between 2008 – 2017. Department of Planning, Monitoring and Evaluation. Available at:

https://www.dpme.gov.za/publications/NIDS%20Wave%205/Poverty%20inequality%20dynamics.pdf [Accessed 16 August 2020]