Measuring human capital in South Africa across socioeconomic subgroups using a latent-variable approach

TJ Friderichs\textsuperscript{1} · F. M. Correa\textsuperscript{2}

Abstract

Human capital is a complex concept to measure given it is an unobserved latent construct. Education is a fundamental dimension of human capital and thus an education-based approach is the one most widely used. However, the international literature recommends a latent-variable approach to measuring human capital. This study thus aims to measure human capital in South Africa, a country experiencing extreme earnings and education inequalities, using a latent-variable approach and the National Income Dynamic Study (NIDS) dataset. The findings are that parental education is associated with the largest amount of variance in (latent) human capital, while the health indicator captures the least variance. Furthermore, the (latent) human capital variable provides a valuable measure to profile the distribution of human capital by socioeconomic subgroups.

Keywords Latent-variable · Human capital · Education · Inequality

JEL codes · I24 · J24 · P36

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”. At the microeconomic level this translates to the HC of each household\textsuperscript{1} (HH) or individual being the most valuable national asset. Thinking about the concept of HC dates back as

\textsuperscript{1} The National Income Dynamics Study, the South African data collection survey used in this study, defines household membership as “spending more than 15 days in the last 12 months at the household and sharing food and resources when staying at that household” (Brophy et al., 2018:17).

TJ Friderichs
t.friderichs@ru.ac.za

\textsuperscript{1} Department of Economics and Economic History, Rhodes University, Makhanda, South Africa

\textsuperscript{2} Department of Statistics, Rhodes University, Makhanda, South Africa
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 how to enhance national wealth accumulation (Folloni and Vittadini, 2010). Controversy shrouded earlier discussions regarding HC as it was believed morally wrong to value humans as resources contributing to wealth (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 microeconomic 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 the innate abilities of individuals 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).

Historically, three standard approaches to measuring HC (Le et al., 2005) have emerged. These are: an income-based measure, originally modeled by Farr (1853); a cost-based measure (Eisner, 1989); and, lastly, various education-based methods (Barro and Lee, 1993 and Woßmann, 2003). More recently a latent-variable approach to measuring HC has been introduced and extended to measure HC in monetary terms (Dagum and Slottje, 2000). This approach was suggested because HC is statistically defined as a multidimensional non-observed latent variable (Dagum and Slottje, 2000). This study thus aims to measure HC availability in South African by applying a latent-variable approach, in a Structural Equation Modelling (SEM) framework, using as reflective indicators education, father’s and mother’s education, school quintile and health. Following the latent HC estimation, the results are compared with the education-based approach, the most widely used to measure HC. Furthermore, considering the persistent inequalities in South Africa, the study examines the distribution of HC, estimated using both measurement approaches, according to different characteristics of the population, namely gender, race, geographical area and income quintile. The value of estimating HC and the distribution thereof, using the different approaches, is thus explored.

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. Thereafter the methodology, including the data, analytical approach and variables used in this paper, are explained. In Sect. 5 the results, including the distribution of the latent HC estimate by subgroup are presented. The final section draws conclusions.

2 Human Capital as a determinant of inequality in South Africa

South Africa has one of the highest levels of wealth and income inequality in the world (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, 2004). 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., 2010 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 education 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 in South Africa 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 socioeconomic status (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. Spaull (2015), for example, states that a child’s future socio-
economic 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., 2014). This suggests that an unequal distribution of income/wealth may negatively impact HC inequality and vice-versa (Chani et al., 2014). This would furthermore explain why in South Africa, or alternate countries with high levels of inequality, talented citizens are unfortunately not likely to reach their full potential, which negatively impacts much needed social and economic development (Keeton, 2014).

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. These education proxies were however noted as misspecifying HC, specifically on a theoretical basis (Wößmann, 2003) and fail to account for the multidimensional nature of HC (Folloni and Vittadini, 2010). The most used proxy, namely average years of schooling, measured using individual workers highest education attainment (Wößmann, 2003), was critiqued for not allowing for returns to education as well as the quality of the education received. 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. Education attainment as a proxy for HC 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 and has been statistically defined as a multidimensional composite variable, it needs to be measured in a latent variable framework (Folloni and Vittadini, 2010). Dagum and Slottje (2000) described household HC as the “multidimensional nonobservable 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. Dagum and Slottje (2000) used 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 as indicators of personal ability, home and social environments. The monetary value of the latent HC 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 Slott-
Vittadini et al. (2003), noted, 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 observed 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. They find that 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 Castello-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 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 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 socioeconomic 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.

4 Method

4.1 Data and sample

The study analysed HC in South Africa, a country with high levels of socioeconomic inequalities. 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 26 776 individuals from 7 300 HHs,
but grew with each wave as HHs expanded, to a total of approximately 37368 individuals and 13719 HHs in wave 5 (Brophy et al., 2018). The most recent, Wave 5, conducted in 2017 and released in 2018, is used in this study (Brophy et al., 2018) to conduct a cross-sectional analysis of South Africa in 2017, with panel weights applied to account for panel attrition (Zizzamia et al., 2018). The two main reasons for using the NIDS dataset were, firstly, the comprehensive nature of the socioeconomic data collected, 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., 2010]). 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 to allow for population estimates (SALDRU, 2017).

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) 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 2 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 the distribution of HC using the education-based and latent-variable approaches, and the HC profile by SES indicators, namely, income, gender, race, and geographical area.

4.2.1 Conceptual measurement model and variables

Different path models were explored to estimate the latent variable HC within a SEM framework and it was identified that the latent variable HC is best estimated in South Africa by applying a reflective CFA measurement model. Coltman et al. (2008) noted that in a measurement model 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). 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 1. 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. 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; and Lovaglio, 2008, 2010). These 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. The variables included in this study are detailed in Table 1.

The measurement model of latent HC (HC_L) can be written as a set of 5 regression equations, as captured in Eq. 1 to 5, or illustrated in the conceptual measurement path model presented in Fig. 1 (StataCorp, 2013).

\[ edu_i = \alpha_i + \beta_i HC_{Li} + \epsilon_i \]  
\[ ftheduj = \alpha_j + \gamma_j HC_{Lj} + \epsilon_j \]  
\[ mthedu_k = \alpha_k + \delta_k HC_{Lk} + \epsilon_k \]  
\[ schoolquin_l = \alpha_l + \rho_l HC_{Ll} + \epsilon_l \]  
\[ health_m = \alpha_m + \tau_m HC_{Lm} + \epsilon_m \]

Table 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 (edu) (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 (health) (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** (schoolquin) (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 (fthedu) (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 (mthedu) (Categorical)** | **Variable for estimating HC_E** |
| 1. Primary or less; 2. Secondary; 3. Matric; 4. Tertiary | Highest years of schooling attainment (Min: 0; Max: 18)* |

Notes: *The NIDS education attainment data was coded as per Branson and Leibbrandt (2013). ** Matric is equivalent to achieving a secondary school leaving certificate, which gives access to tertiary education, namely International Standard Classification of Education (ISCED) levels 5 or 6. *** Schools in South Africa are classified into schools from most poor to least poor. Schools receive money from government according to Quintiles, with Quintile 1 schools receiving the highest allocation per learner, while Quintile 5 receives the lowest.
Where $HC_{Li}$ is estimated latent HC; $edu_i$ the education; $fthewdu_j$ the father’s education; $mthewdu_k$ the mother’s education; $schoolquint_l$ the school quintile; $health_m$ health; $\alpha$ is the intercept associated with the latent variable and $\beta_i$, $\gamma_j$, $\delta_k$, $\rho_l$, and $\tau_m$ are the path coefficient for education, father’s education, mother’s education, school quintile and health respectively and $\varepsilon$ the error term associated with the respective equations. The path diagram is illustrated in Fig. 1.

Equation 1 to 5 are used to estimate the path coefficients between the specified unobserved latent ($HC_{Li}$) and observed indicators in the data. Thereafter $HC_{Li}$ per observation is predicted ($\hat{HC}_{Li}$) to analyse the distribution of HC in South Africa.

It should be noted that Fig. 1 does not imply causality, but by accounting for the variance and covariance of the observed variables (Jarvis et al., 2003) the reflective CFA model estimates how well the specified observed indicators represent $HC_{Li}$ (Bagozzi, 2011).

With regards to the observed indicators in the model, 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). 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). Becker (1964:3) included health in the definition of HC, stating that HC is the “knowledge, information, ideas, skills, and health of individuals”. Empirical evidence from South Africa has found that higher education attainment and better health have positive impacts on socioeconomic wellbeing (van Broekhui-
Measuring human capital in South Africa across socioeconomic factors (Bound, 1991; 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 (Ardington & 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.

Following the estimation of the CFA measurement model, using the statistical programme STATA (sem package), the goodness of fit analysis and confirmation of the measurement model, HC is predicted for further analysis (Coltman et al., 2008 and StataCorp, 2013). The results from the reflective model will highlight potential areas requiring additional investments to reduce HC and productivity gaps in South Africa.

i) 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 are presented in Table 2. Included in the exploratory data analysis, a log-linear earnings function, specified in Eq. 6, 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 partial regression coefficients from the log-linear earnings function specified in Eq. 6 are presented in Table 2.

The log-linear earnings function is:

$$\log Y_{ijklm} = \alpha + \beta_i \text{edu}_i + \gamma_j \text{ftedhu}_j + \delta_k \text{mthedu}_{Lk} + \rho_l \text{schoolquin}_l + \tau_m \text{health}_m + \epsilon_{ijklm} \tag{6}$$

Where $\log Y$ is the log of net monthly labour income per capita in ZAR, and \text{edu}, \text{ftedhu}, \text{mthedu}, \text{schoolquin}, and \text{health} are the observed study variables and $\beta_i$, $\gamma_j$, $\delta_k$, $\rho_l$, and $\tau_m$ are the respective coefficients being $\epsilon$ is a vector of residuals per observation (error term). Eq. 6 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 Broekhuiizen; 2011; Branson and Leibbrandt; 2013). The partial regression coefficients 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 with the results discussed in Sect. 5.2. 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

---

2 Eq. 6 is included for the purpose of exploratory data analysis. Estimating the partial regression coefficients is not the main aim of the study, which is to estimate the latent variable HC using a CFA measurement model. The partial regression coefficients are presented with the descriptive statistics in Table 2.
### Table 2 Exploratory data analysis of study variables: Sample data descriptive statistics and partial regression coefficients

| Variable                                      | Observations and descriptive statistics (Post-stratified weights applied) | Estimated partial regression coefficients*. $R^2 = 0.4471$ |
|-----------------------------------------------|---------------------------------------------------------------------------|----------------------------------------------------------|
| $Y$ (individual monthly labour income)        | Obs 5 321 615; Mean 9 370.77; Std dev 14 745.83                           | Dependant variable                                        |
| $\log Y$ (log labour income per month)        | Obs 5 321 615 Mean 8.434; Std dev 1.226                                  |                                                          |
| **Education attainment**                      | 5 321 615                                                                 |                                                          |
| 1 Primary or less                             | 9.93% Std error (0.0082)                                                  | Reference group                                           |
| 2 Secondary                                   | 34.14% Std error (0.0148)                                                 | 0.30** 0.00                                              |
| 3 Matric                                      | 22.01% Std error (0.0135)                                                 | 0.67** 0.00                                              |
| 4 Some tertiary                               | 22.73% Std error (0.0137)                                                 | 1.32** 0.00                                              |
| 5 Bachelor degree or more                     | 11.19% Std error (0.0087)                                                 | 1.98** 0.00                                              |
| **Self-perceived health status**              | 5 321 615                                                                 |                                                          |
| 1 Poor                                        | 1.78% Std error (0.0041)                                                  | Reference group                                           |
| 2 Fair                                        | 6.23% Std error (0.0077)                                                  | -0.06 0.74                                               |
| 3 Good                                        | 24.17% Std error (0.0131)                                                 | 0.09 0.55                                                |
| 4 Very good                                   | 32.81% Std error (0.0146)                                                 | 0.04 0.76                                                |
| 5 Excellent                                   | 35.01% Std error (0.0151)                                                 | 0.12 0.43                                                |
| **School quintile**                           | 5 321 615                                                                 |                                                          |
| 1 No-fee school                               | 24.59% Std error (0.014)                                                  | Reference group                                           |
| 2 No-fee school                               | 20.26% Std error (0.0119)                                                 | -0.01 0.93                                               |
| 3 No-fee school                               | 26.16% Std error (0.0135)                                                 | -0.03 0.72                                               |
| 4 Low-fee School                              | 10.94% Std error (0.0098)                                                 | 0.11 0.43                                                |
| 5 Fee charging school (private and public)    | 18.05% Std error (0.0123)                                                 | 0.16 0.23                                                |
| **Mothers education attainment**              | 5 321 615                                                                 |                                                          |
| 1 Primary or less                             | 57.16% Std error (0.0156)                                                 | Reference group                                           |
| 2 Secondary                                   | 21.70% Std error (0.0126)                                                 | 0.07 0.43                                                |
| 3 Matric                                      | 11.11% Std error (0.0105)                                                 | -0.04 0.82                                               |
| 4 Tertiary                                    | 10.03% Std error (0.0101)                                                 | 0.13 0.34                                                |
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).

ii) 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 (Fig. 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 indi-
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 ≥ 0.90 (Bentler, 1990); TFI values ≥ 0.90 (Hu and Bentler, 1999); SRMR values ≤ 0.06 (Hu and Bentler, 1999) and RMSEA values ≤ 0.08 (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, which is factor score 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.

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 Fig. 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 4.

Secondly, an alternative DRT, namely 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 dataset, is used to predict $HC_p$. The factor weights per component are presented in Table 5.

### Table 3: CFA measurement model: Standardized coefficient estimates, goodness of fit results and predicted HC estimates

| Effect                  | Std. coefficient | S.E. | $R^2$ | P-value |
|-------------------------|------------------|------|-------|---------|
| $HC_L$ ~ Education      | 0.541            | 0.024| 0.30  | 0.000   |
|                         | 0.823            | 0.021| 0.68  | 0.000   |
| Mother’s education      | 0.831            | 0.021| 0.69  | 0.000   |
| Father’s education      | 0.158            | 0.034| 0.02  | 0.000   |
| Health                  | 0.513            | 0.033| 0.24  | 0.000   |
| School quintile         | Goodness of      | Benchmark | Result | Good fit |
|                         | fit results*     |       |       |         |
| CD                      | 1.00             | 0.837| Yes   |         |
| SMRM                    | ≤ 0.06           | 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

---

3 $HC_L$ 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.

4 The factor scores are estimated by means of a linear regression, using the fitted model’s mean vector and variance matrix.
iii) Analysing the profile of HC by SES indicators.

Firstly, to map the concentration distribution of HC, Kernel density (k-density) curves for $HC_h$, $HC_f$, $HC_p$, $HC_p$, and $HC_E$ are illustrated in Fig. 2. Thereafter, a histogram (in Fig. 3) is presented to compare the frequency distribution of $HC_L$ and $HC_E$. Finally, Tables 6 and 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 which

Table 4 Adjusted CFA measurement models: Standardized coefficient estimates and goodness of fit results

| Effect          | $HC_h$ (no health) | $HC_f$ (no father’s educational attainment) |
|-----------------|--------------------|---------------------------------------------|
| Education       | 0.538              | 0.558                                       |
| Mother’s education | 0.821             | 0.795                                       |
| Fathers education | 0.834             | 0.022                                       |
| Health          | 0.513              | 0.183                                       |
| School quintile | 0.033              | 0.036                                       |

Goodness of fit results

| Benchmark       | $HC_h$ | $HC_f$ | $HC_p$ | $HC_p$ | $HC_E$ |
|-----------------|--------|--------|--------|--------|--------|
| Post-stratified weights applied* |        |        |        |        |        |
| CD              | Close to 1 | 0.837 | Yes | 0.721 | Yes |
| SMRM            | ≤0.06   | 0.002 | Yes | 0.016 | Yes |

No weights applied**

| CFI             | ≥0.90  | 0.999 | Yes | 0.987 | Yes |
| TFI             | ≥0.90  | 0.999 | Yes | 0.962 | Yes |
| SMRM            | ≤0.06  | 0.007 | Yes | 0.021 | Yes |
| RMSEA           | ≤0.08  | 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

Table 5 Factor weights for observed indicators

| Variables                  | Principal component 1 (PC1) | Principal component 2 (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                     | 0.970373                    |

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

$^5$ K-density curves illustrate the density of a variable from the observations of the variable.
T. Friderichs, F. M. Correa 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.

5 Results and discussion

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 2. Furthermore, the partial regression coefficients estimated from Eq. 6 are presented.

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

![K-density curves: HC distribution results](source: NIDS wave 5; authors’ calculations. Post-stratified weights applied)
improved over the generations, with over 50% of parents obtaining primary schooling or less, compared to 9.93% of the earlier 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 Bachelors degree or more) remain low, at 11.19% of employed individuals. This is concerning, given that returns to education are high and increasing for degree graduates in South Africa (van Broekhuizen; 2011; 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 increased returns to higher levels of education is supported by the partial regression coefficients presented in Table 2. Having a Bachelors degree or more is correlated with earnings 198% higher in comparison to those with primary schooling or less. 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; and Whites, Asian/Indians and Coloureds are expected to earn 60%, 57% and 18% respectively more than Africans. Comparing males and females 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 Eq. 6 and the sample studied.
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 ($\chi^2 = 3204.67$, $p = 0.0000$) and the KMO test coefficient was 0.739 and thus the factor analysis could be conducted.

5.2 Confirmatory factor analysis (CFA) measurement model

The CFA measurement model has 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. 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\textsuperscript{6}) on labour force participation of between 20% and 33%.

\textsuperscript{6} The econometric techniques used included instrumental variables, censored quantile regression and Blinder-Oaxaca decomposition.

| Table 6 | Human capital (latent) quintiles by gender, race, geographical area and income quintile |
|---------|-------------------------------------------------------------------------------------|
|         | 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        | 12.0 (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 alongside each percentage estimate. $HC_L$ 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.
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 $H_{CL}$ construct. Providing a broader range of goodness of fit results, the unweighted estimates, illustrated in Appendix 2, confirm the goodness of fit of the model and the estimated path-coefficients. 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 $H_{CL}$ construct respectively. Appendix 2 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 & Green, 2010). It is therefore 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 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 $H_{CL}$ estimates, with only negligible differences.

The factor scores presented in Table 5, estimated using PCA, indicate comparable variance estimations identified using the CFA measurement model to predict $H_{CL}$, as indicated in Table 3. The results in Tables 3 and 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 $H_{C_p}$. Furthermore, the proportion of variance explained by PC1 is double that of PC2.

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. $H_{C_L}$ per individual is therefore estimated for analysis by gender, race, geographical area and income in South Africa.

5.3 Human capital frequency distribution by subgroup in South Africa

From 2770 observations the predicted $H_{C_L}$ 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 $H_{C_L}$ is presented in Fig. 2 using k-density curves. Numerous individuals are estimated as having low levels of $H_{C_L}$. The sensitivity analysis illustrates the robustness of the results of the concentration of $H_{C_L}$ 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 (see Table 3); thus the variance of $H_{C_L}$ 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.

The histogram presented in Fig. 3 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 $H_{C_L}$, however moderate levels of $H_{C_E}$.

The distribution profile of $H_{C_L}$ and of $H_{C_E}$ by race, gender, geographical area and income quintile are presented in Tables 6 and 7, respectively. In Table 6 the $H_{C_L}$ quintiles are calculated by dividing the population into five groups of equal number according to the value of their $H_{C_L}$. The findings illustrate that the individuals in the sample with the top 20% of $H_{C_L}$ are most commonly in income quintile 5, White, male and resident in urban areas. The subgroup profile of $H_{C_E}$ is presented in Table 7 considering five education categories. Individuals in the top $H_{C_E}$ category are predominantly income quintile 5 earners, White, female and resident in urban areas. It should be noted that $H_{C_L}$ and $H_{C_E}$ estimates are not directly comparable by subgroup using quintile distributions given data limitations. Thus, the subgroup distribution of $H_{C_L}$ is presented by quintile and of $H_{C_E}$ 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).

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 $H_{C_L}$. In contrast, 35.3% of rural dwellers are in the lowest $H_{C_L}$ quintile (compared with only 15%). The $H_{C_E}$ categories confirm the dominance of higher levels of HC in urban...
areas, however, HC_L illustrates larger differences between urban and rural areas when comparing quintile 5 levels of HC_L with the highest HC_E category, namely Bachelors degree or more, as evident in Tables 6 and 7.

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 Bachelors 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 earnings partial regression coefficient for gender illustrates 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 & 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 Whites. 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 HC in explaining income distribution in South Africa. 66.8% of income quintile 1 and 56.7% of income quintile 2 have the two 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 returns to education 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 cannot be accurately presented when using the HC_E categories. 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 the 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 outcomes is likely 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 \(H_{CL}\) and earnings, further research on the returns to \(H_{CL}\) is recommended.

6 Conclusions

Human capital 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 human capital, mainly because there is extensive data available. However, a latent-variable approach has been proposed, since human capital is in fact a multidimensional unobserved construct and the potential for proxy measurement errors. The authors employed a reflective confirmatory factor analysis measurement model to estimate the latent construct human capital 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 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 human capital theory, and sensitivity analyses were conducted to analyse the robustness of the results. Human capital profiles by (latent) human capital quintiles and (education) human capital categories are presented.

Valuable insights are gained from estimating human capital in a confirmatory factor analysis measurement model and presenting the (latent) human capital and (education) human capital profiles by socioeconomic status indicators. Firstly, it was found that in South Africa parental education is associated with the largest amount of variance in human capital, while the health indicator captures the least. Human capital (latent) is noted as illustrating a greater frequency distribution in lower levels of human capital in comparison to (education) human capital. The differences are highlighted because of the inclusion of parental education and school quality in (latent) human capital, which captures the socioeconomic status of HHs. Furthermore, when studying human capital frequency distribution by population subgroup in South Africa, the (latent) human capital variable allows for analysis by population quintiles, which human capital measured using an education-based approach does not. It is identified that the individuals in the sample with the top 20% of (latent) human capital are most commonly income quintile 5 earners, White, male and resident in urban areas. The subgroup profile of (education) human capital, presented using five education categories, estimates individuals in the top (education) human capital category to be predominantly income quintile 5 earners, White, female and resident in urban areas. However, although the results do not appear to differ significantly, the (latent) human capital measure accentuates distribution differences. For example, when comparing racial distributional differences, the (latent) human capital estimates emphasise the unequal development of HC between Whites and Africans under the apartheid-era continues, which although evident when studying (education) human capital are not as accentuated.
Thus, it is suggested that without improvements in schooling quality, valuable education attainment and parental education for those historically disadvantaged and subject to poor service delivery, will remain low and thus inequalities in human capital 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 human capital, the process of measuring human capital using a latent-variable approach in this study highlights why the education-based approach to measuring human capital persists in South Africa. The reasons include data availability, simplicity of the approach and that the education-based human capital variable presents expected distributional differences between subgroups, with income quintile 5 earners, Whites, and residents in urban areas experiencing higher levels of human capital in comparison to lower income quintile earners, Africans, and residents in rural areas. The latent-variable approach however allows for population quintile analysis and multiple indicators to be incorporated so as to account for the intergenerational transfer of poverty and inequality and a schooling system which perpetuates this cycle. For future research, policymakers from additional countries could benefit from exploring a structural equation modelling framework to analyse human capital as a latent variable. Furthermore, research focused on understanding health as an indicator of human capital is also recommended.

7 Appendix

7.1 Appendix 1

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  |               |      |                   |        |
### 7.2 Appendix 2

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      | ≥0.90     | 0.994  | Yes      |
| TFI      | ≥0.90     | 0.987  | Yes      |
| SMRM     | ≤0.06     | 0.019  | Yes      |
| RMSEA    | ≤0.08     | 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

### Declarations

**Competing interests** The authors declare they have no competing interests related to this research.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.
References

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

ARDINGTON, C., & 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

BABYAK, M. A., & 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.0b013e3181ded38a [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., & LEE, J. W. (1993). International comparisons of educational attainment. *Journal of Monetary Economics, 32*(3), 363–394

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

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

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

BOWLES, S., & 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., & LEIBBRANDT, M. (2013). M. 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., & 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: https://doi.org/10.1787/5k4e0vvbvv0q-en [Accessed 10 January 2020]

BREEN, R., & 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). *Employment and Earnings in South Africa – A Cohort Analysis*. SALDRU Working Paper 105. Cape Town: Southern Africa Labour and Development Research Unit

BROPHY, T., BRANSON, N., ARDINGTON, C., & GASEALAHWE, B. (2012). *National Income Dynamics Study Panel User Manual*. National Income Dynamics Study. Available at: http://www.nids.uct.ac.za/images/documents/20180831-NIDS-W5PanelUserManual-V1.0.pdf [Accessed 10 January 2019]

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

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

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

CHATTERJEE, A., CZAJKA, L., & 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]

CHIAPPA, R., & 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]

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

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

EISNER, R. (1989). The total incomes system of accounts. *Survey of Current Business, 65*(1), 24–48

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

FOLLONI, G., & VITTADINII, G. (2010). Human Capital Measurement: A Survey. *Journal of Economic Surveys, 24*(2), 248–279

GROSSMAN, M. (2000). The Human Capital Model. *Handbook of Health Economics, 1*, 347–408
HUNDENBORN, J., LIEBBRANDT, L., & 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., & 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

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

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

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

LAVERDE-ROJAS, H., CORREA J.C., JAFFE, K. & 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. (2005). *Measurements of Human Capital: A review of the literature*. New Zealand Treasury Working Paper 05/10. Wellington, New Zealand: New Zealand Treasury

LEIBBRANDT, M., WOOLARD, I., MCEWEN, H., & 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., & WOOLARD, I. (2012). Describing and decomposing post-apartheid income inequality in South Africa. *Development Southern Africa, 29*(1), 19–34

LOEHLIN, J. C., & 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

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

MAGADLA, S., LEIBBRANDT, M., & 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

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

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

MENCER, J. (1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy, 66*(4), 281–302

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

MOOI, E., SARSTEDT, M., & 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]

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]

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

ORGANISATION for ECONOMIC CO-OPERATION and DEVELOPMENT (OECD) (2001). *The well-being of nations: The role of human and social capital*. Paris: Centre for educational research and innovation

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

PELLICER, M., & 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., & SULLIVAN, J. (2003). *Making sense of factor analysis*. Thousand Oaks: Sage Publications, Inc.

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

SUHRCKE, M., MCKEE, M., STUCKLER, D., SAUTO ARCE, R., TSOLLOVA, S., & 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]

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 N. Spaull, & D. Jansen (Eds.), (eds), *South African Schooling: The Engine of Inequality* (pp. 1–24). Switzerland: Springer Nature

REPUBLIC OF SOUTH AFRICA (1996). *South African Schools Act 84 of 1996*. Available at: https://www.gov.za/sites/default/files/gcis_document/201409/act84of1996.pdf [Accessed 19 March 2022]

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]

UNITED NATIONS (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

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

VAN DER BERG, S., & 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), 546–572

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 https://doi.org/10.1080/0376835X.2013.871196 [Accessed 5 January 2020]

VITTADINI, G., DAGUM, C., LOVAGLIO, P. G., & 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. & 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: University of Western Cape, Department of Economics

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

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

WORLD BANK (2019). *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]

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.