A case study variate analysis of low-income learner progression in South Africa's high school system

by

Jemma Ann Visser

Thesis presented in partial fulfilment of the requirements for the degree of
Master of Engineering (Industrial Engineering)
in the Faculty of Engineering at Stellenbosch University

Supervisor: Prof Imke de Kock
Co-supervisor: Dr Lieschen Venter

December 2025

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: December 2025

Abstract

This study investigates the determinants of academic progression among low-income learners within two contrasting South African educational contexts: Low-income schools and higher-income schools. Drawing on data from the 2022 General Household Survey, the analysis employs factor analysis, multivariate regression, structural equation modelling, and machine learning techniques to identify eight latent constructs that characterise learners from lower socio-economic backgrounds, namely, family structure, learner health, supported retention, food security, welfare income, sanitation, energy, and environmental conditions.

Although learners across both contexts share similar socio-economic disadvantages, the influence and function of these factors diverge considerably. Learners attending schools with lower socio-economic status tend to benefit from greater institutional support and increased access to welfare income, yet they often face more precarious household conditions. In contrast, learners enrolled in higher status schools are generally situated in more stable home environments but receive comparatively less institutional assistance. The analysis reveals that regular school attendance, age appropriate grade placement, access to school-provided meals, and welfare income are consistently critical for academic progression across both contexts, though their relative significance varies.

To substantiate these findings, the analytical approach was applied to an independent high school case study. This application confirmed that, once foundational needs are adequately met, higher-order factors such as paternal involvement, future-oriented optimism, and intrinsic motivation become increasingly influential in determining academic success. These insights informed the development an educational hierarchy of needs decision-support model, which outlines five sequential levels essential for the academic development of low-income learners. Level 1 addresses foundational needs, including regular school attendance and age-appropriate grade placement. Level 2 emphasises the necessity of access to school meals to ensure adequate nutrition. Level 3 focuses on financial stability, particularly the receipt of welfare income such as the child support grant. Level 4 centers on the importance of family support structures, with a particular emphasis on paternal involvement. Level 5 culminates in mental well-being, marked by self-motivation and a positive outlook on the future.

The decision-support model is grounded in the principle that educational progression must follow a bottom-up trajectory, whereby higher-level development is contingent upon the fulfilment of lower-level needs. For instance, strong psychological well-being (Level 5) cannot compensate for irregular school attendance (Level 1) or nutritional deficiencies (Level 2). As such, this hierarchical decision-support framework offers a structured and systematic approach to identifying and addressing the most pressing barriers to educational success for low-income learners within South Africa's bifurcated education system, where profound disparities persist along socio-economic lines.

Opsomming

Hierdie studie ondersoek die onderliggende faktore wat akademiese vordering onder lae-inkomste leerders beïnvloed, binne die twee kontrasterende Suid-Afrikaanse onderwyskontekste van lae-inkomsteskole en hoër-inkomsteskole. Die analise steun op data uit die 2022-weergawe van die Algemene Huishoudelike Opname en gebruik metodologiese tegnieke soos faktoranalise, meerveranderlikeregressie-analise, strukturele vergelykingsmodellering en masjienleer. Hierdie benadering het agt latente faktore geïdentifiseer wat leerders uit lae sosio-ekonomiese agtergronde kenmerk, naamlik gesinstruktuur, leerdergesondheid, ondersteunende behoud, voedselsekerheid, welsyninkomste, basiese higiënefasiliteite, toegang tot energie en omgewingsomstandighede.

Hoewel leerders in beide kontekste soortgelyke sosio-ekonomiese uitdagings ervaar, verskil die aard en invloed van hierdie faktore beduidend. Leerders in skole met 'n laer sosio-ekonomiese status geniet dikwels groter institusionele ondersteuning en beter toegang tot welsynshulp, maar leef in meer onstabiele huishoudings. Daarteenoor leef leerders in hoërstatuskole gewoonlik in stabieler huishoudelike omgewings, maar ontvang minder eksterne ondersteuning. Die ontleding toon dat gereelde skoolbywoning, ouderdomsgepaste graadplasings, toegang tot skoolvoeding en welsynsinkomste sleutelvoorspellers van akademiese vordering is in albei kontekste, hoewel hul relatiewe gewig verskil.

Ter bevestiging van hierdie bevindings is dieselfde analitiese raamwerk toegepas op 'n onafhanklike hoërskool-gevallestudie. Hierdie toepassing het getoon dat, sodra basiese behoeftes aangespreek is, hoër-orde faktore soos vaderlike betrokkenheid, toekomsgerigte optimisme en intrinsieke motivering al hoe meer bepalend word vir leerderprestasie. Hierdie insigte het aanleiding gegee tot die ontwikkeling van 'n hiërargiese besluitsteunmodel vir opvoedkundige ingryping, wat vyf opeenvolgende vlakke voorstel wat noodsaaklik is vir die akademiese ontwikkeling van leerders uit lae-inkomste huishoudings.

Die grondvlak fokus op fundamentele behoeftes soos gereelde skoolbywoning en graadplasings. Die tweede vlak beklemtoon voldoende voeding deur toegang tot skoolvoedsel. Die derde vlak lê klem op finansiële sekuriteit, insluitend welsynsvoordele soos die kindondersteuningstoelae. Die vierde vlak behels gesinssteun, met spesifieke fokus op vaderlike betrokkenheid. Die vyfde vlak beklemtoon geestesgesondheid, gekenmerk deur selfmotivering en 'n positiewe lewensuitkyk.

Die model is gegrond op die beginsel dat ontwikkeling volgens 'n opwaartse trajek verloop. Hoërvlakvaardighede kan slegs ontwikkel word indien laervlakbehoeftes vervul is. Byvoorbeeld, geestelike welstand (vlak vyf) kan nie 'n gebrek aan skoolbywoning (vlak een) of voeding (vlak twee) kompenseer nie. Hierdie raamwerk bied dus 'n gestruktureerde metode om sleutelhindernisse tot akademiese vordering in Suid-Afrika se verdeelde onderwyslandskap te identifiseer en aan te spreek.

Acknowledgements

The author wishes to acknowledge the following people for their various contributions towards the completion of this work:

- First and foremost, I attribute all glory and honor to our Lord and Savior, Jesus Christ, for the talents and opportunities bestowed upon me. I express gratitude for the perseverance and courage with which He has enriched me, enabling the successful completion of this project.
- My husband and best friend, Hendrik Visser, for his steadfast support, encouragement, and belief in my work. Your presence throughout this research journey has been a constant source of strength, and I am sincerely grateful for your unwavering cheer and companionship.y my side and always been there to cheer me on.
- My parents, Deon and Helen Bührs, for their unwavering motivation, encouragement, and support throughout research journey. I am deeply thankful for the emotional, physical, spiritual, and financial guidance generously bestowed upon me.
- My supervisors, Dr Lieschen Venter and Prof Imke de Kock, for your invaluable time and willingness to collaborate with me. Your consistent guidance in the right direction is profoundly appreciated.

Contents

List of Acronyms

ABM	Agent-based modelling
CAPS	Curriculum and Assessment Policy Statement
CFA	Confirmatory factor analysis
CFI	Comparative fit index
CLD	Causal loop diagramme
DBE	Department of Basic Education
DHET	Department of Higher Education and Training
EFA	Exploratory factor analysis
FA	Factor analysis
GHS	General Household Survey
HSM	High School Model
KMO-MSA	Kaiser-Meyer-Olkin measure of sampling adequacy
LCA	Latent class analysis
LQ13	Learner Quintile 1-3
ML	Machine learning
MRA	Multivariate regression analysis
NSES	National School Effectiveness Study
NSC	National Senior Certificate
PCA	Principle component analysis
PIRLS	Progress in International Reading and Literacy Studies
PSM	Primary School Model
ReSEP	Research on Socio-Economic Policy
RMSEA	Root mean square error of approximation
SACMEQ	Southern and East Africa Consortium for Monitoring Educational Quality
SD	System dynamics
SEM	Structural equation modelling
SES	Socio-economic status
SMOTE	Synthetic minority over-sampling technique
SQ13	School Quintile 1-3
SQ45	School Quintile 4-5
SRMR	Standardised root mean square residual
STEP	Systems Thinking for Education Policy
TIMSS	Trends in International Mathematics and Science Study

List of Figures

List of Tables

CHAPTER 1

Introduction

Contents

1.1 Background

Situated at the southern tip of Africa, South Africa is marked by the historical imprint of apartheid, an institutionalised system of racial segregation. The struggle against apartheid, pioneered by leaders such as Nelson Mandela, culminated in the nation's inaugural democratic elections in 1994 [85]. Despite comprising a well-established industrial sector, abundant mineral wealth, and a successful agricultural domain, the nation confronts a pronounced socio-economic crisis, marked by an elevated unemployment rate of 31.9% as of 2023 [90]. This crisis is further exemplified by the fact that in 2023, 34.2% of the youth (*i.e.*, those aged 15 to 24 years) were neither employed nor engaged in education or training [60]. Such figures portend significant implications for South Africa's future, spanning economic prospects, social development, and the daunting challenges in addressing youth unemployment and educational attainment disparities.

Reflecting on the words of the influential figure in the civil rights movement, Malcolm X,

> "Education is the passport to the future, for tomorrow belongs to those who prepare for it today."

emphasises the pivotal role of education in shaping the trajectory of a country. Improving the education system is of paramount importance as it serves to equip citizens with the required knowledge and skills that are imperative for their meaningful integration as productive contributors to the South African society [9]. This, in turn, augments the competitiveness of South Africa within the global economy, thereby presenting a potent antidote to the prevailing socio-economic crises faced within the country.

1.1.1 Introduction to the South African education system

The governance of the South African education system is overseen by the Department of Higher Education and Training (DHET) [89] and the Department of Basic Education (DBE) [88]. The DHET directs vocational training and higher education, emphasising post-school education and training, while the DBE is responsible for primary and secondary school education. The DBE oversees public schools, independent schools, early childhood development centres, and special needs education facilities up to Grade 12. DBE's vision is to provide equal access to education for all within South Africa.

The public education system comprises four distinct phases through which learners progress from entry to the successful completion of the National Senior Certificate (NSC) [58]. A learner enters the system during the foundation phase at the age of five years in Grade R (the reception year) and remains until Grade 3. The intermediate phase begins in Grade 4 at the age of nine years and concludes in Grade 6. The senior phase commences at twelve years of age in Grade 7 and ends in Grade 9. The final phase is Further Education and Training (FET), which begins at age fifteen in Grade 10 and culminates in Grade 12 (the matriculation year) after successful completion of the NSC examination. This progression reflects the prescribed grade–age trajectory. However, due to factors such as early or late school entry and grade repetition, actual learner progression may deviate from this pattern. The policy objective is to encourage adherence to this progression path.

Generally, the public education system is organised into the primary school phase (*i.e.*, Grades 1 to 7) and the high school phase (*i.e.*, Grades 8 to 12). According to the South African Schools Act of 1996 [91], children are obligated to attend school from the beginning of the year in which they turn seven years old until the end of the year in which they turn fifteen.

The South African education system is characterised by two the distinct and parallel education systems, namely the public system and the independent system [58]. Public schools, which are government-owned and funded, often offer a more economically accessible option for learners. In contrast, independent schools are privately owned, funded, and operated. In 2022, out of the 13.4 million learners in South Africa, 94.5% were registered within the public education system, with the remaining 5% attending independent educational institutions [28].

For learners opting out of the mainstream schooling system, attending a School of Skills presents a viable option [102]. These institutions cater to students struggling in conventional settings, offering practical skills aligned with labour market demands. Students receive academic and technical training tailored to their abilities and interests, preparing them for vocational training or employment. Skills offered range from automotive repair to needlework and housekeeping, addressing diverse needs within the educational system. Thus, Schools of Skills play a crucial role in facilitating educational attainment and future prospects for learners with varied requirements. By equipping students with practical skills and vocational training, Schools of Skills enable them to make valuable contributions to the economy through employment and entrepreneurship, thereby fostering economic growth and sustainability.

In an attempt that all learners within public schools in South Africa receive equal education, the DBE provides schools with the national Curriculum and Assessment Policy Statement (CAPS) [86]. CAPS provides the policy for learning and teaching from Grade R to Grade 12. Each specific subject is governed by a CAPS document, encompassing an introductory section outlining the objectives of the curriculum, relevant teachings and expected learning outcomes for the subject, the requisite content to be covered, and the methodology and policy for evaluating acquired knowledge. In essence, CAPS serves as a framework aimed at preparing learners for the NSC, which is undertaken at the culmination of the matriculation year (which is the final year of secondary school), commencing from the first day of formal education in Grade R [87].

The public school system is subdivided into quintiles based on socio-economic status (SES)[1] [23]. Quintile 1 to 3 schools are indicative of lower SES-ranked schools, which benefit from a more substantial percentage of the annual education budget provided by the government. Schools falling within this range are designated as no-fee schools, where learners are not required to pay school fees [24]. Conversely, Quintile 4 and 5 schools serve learners from a higher SES ranking and consequently receive a more limited percentage of funding from the government. These schools possess the autonomy to levy school fees and are thus fee-paying institutions. Public schools operate on a zoning basis, prioritising learners residing within a specified kilometre radius from the school [35].

1.1.2 The state of the South African public high school system

The current South African education system faces significant challenges, manifesting in declining academic performance and persistent disparities among learners. These challenges expose the urgent need for comprehensive reforms to address issues of enrolment patterns, gender disparities, and socio-economic inequalities within the education system.

Various assessments are employed internationally to determine educational achievement. The Progress in International Reading and Literacy Study (PIRLS) [46] is a framework that assesses a learner's reading achievement in Grade 4. In this grade, a learner transitions from learning to read to reading to learn; thus, PIRLS evaluates a learner's ability to read accurately and for meaning. According to the 2016 PIRLS report, slightly more than one in every four South African learners in Grade 4 exhibited the ability to read coherently.

Similarly, the Trends in International Mathematics and Science Study (TIMSS) [47] assesses learners' mathematical and scientific knowledge and understanding. The 2019 TIMSS findings indicated that one in every three Grade 9 South African learners was overage for their grade, suggesting prior grade repetition. Moreover, only 41% of Grade 9 learners in South Africa reached the low benchmark in mathematics, signifying an ability to demonstrate some proficiency in whole numbers and basic graphing concepts. Similarly, only 36% of learners reached the low benchmark in science, which encompasses basic knowledge in biology, physical sciences, and earth sciences [72].

The Research on Socio-Economic Policy (ReSEP) group at Stellenbosch University focuses on issues related to poverty, education, social mobility, income distribution, and social policy [73]. In 2023, Wills and Qvist [103], two researchers at ReSEP, released a report summarising the impact of the COVID-19 pandemic on South Africa's education system, drawing from administrative records and household surveys. They revealed a nuanced landscape characterised by shifts in

[1] SES is determined by a combination of social and economic factors, including occupation, income, education level, and residential location [7].

enrolment patterns, persistent gender and socio-economic disparities, and cautious optimism regarding post-pandemic recovery.

Initially, there were concerns about increasing attrition and repetition rates in South African schools as a result of the pandemic. However, administrative records and household surveys indicate that although non-participation and extended absenteeism increased, realised attrition and repetition rates actually decreased. Specifically, in 2020, prior to the pandemic, approximately 11% of adolescents aged fifteen to nineteen years were neither enrolled in school nor had completed secondary education. This figure declined to 8.6% in 2021 and rose slightly to 10.2% by 2023. Furthermore, the Grade 10 repetition rates in 2019 and 2020 were 31% and 17% respectively, and a shift in enrolment patterns was observed [36]. The number of learners enrolled in secondary education increased, partially due to lenient progression requirements and the lack of economic alternatives outside of school.

A 2022 report by Van der Berg *et al.* [98] noted a concerning trend in South African basic education, revealing a decline in academic performance among Grade 3, 6, and 9 learners. To attain a passing score, learners must achieve a minimum mark of 30%. Specifically, the proportion of Grade 3 learners passing a language test decreased from 68% in 2019 to 59% in 2021. Similarly, the percentage of Grade 6 learners achieving a passing score decreased from 85% in 2019 to 76% in 2021. This decline in performance was also evident in systemic tests for both language and mathematics across Grades 3, 6, and 9, with average scores in 2021 falling below those recorded in 2019. In Grade 3, the average language scores decreased from 42.4% in 2019 to 38.7% in 2021, while the average mathematics scores decreased from 59.5% to 50.7% during the same period. For Grade 6, the average language scores declined from 50.5% in 2019 to 45.0% in 2021, and the average mathematics scores fell from 55.7% to 47.3%. Similarly, in Grade 9, the average language scores decreased from 59.1% in 2019 to 56.2% in 2021, and the average mathematics scores dropped from 37.7% to 31.5%. This casts a shadow of doubt on any optimism that the improved grade progression rates shown by Wills and Qvist could have evoked.

The Reading Panel 2030 [1] constitutes a group of experts tasked with evaluating and analysing the prevailing education system to ascertain the underlying causes behind the inability of children to read with meaning. Their evaluation encompasses the review of governmental initiatives and strategies, with the objective of securing adequate political attention and will to address the ongoing reading crisis. The overarching goal of the Reading Panel 2030 is to achieve literacy among all South African children by the year 2030.

In the South African education system, Spaull [92] discussed bimodality and the existence of two distinct levels of academic performance among learners. He showed that there were two observable groups with one performing at an acceptable level and the other struggling academically. These differences can be seen across different grade levels, subjects, socio-economic backgrounds, and types of schools.

Spaull offered three illustrative examples. Firstly, findings from the third Southern and East Africa Consortium for Monitoring Educational Quality (SACMEQ III) [62] revealed a striking bimodal distribution in reading scores. One mode of learners (from families within the wealthiest 25% of South African families) achieved an approximate reading score of 80%, while the other mode (of learners from families within the poorest 75% of South African families) achieved merely around 40%.

Additionally, data from the PIRLS [63] conducted in 2006 provides further evidence of bimodality in South African learners. Specifically, one mode of learners (from English or Afrikaans speaking schools) attained a reading proficiency level of approximately 75%, while the other mode (of learners from an African language speaking school) achieved a significantly lower score of around 25%.

Finally, insights from the National School Effectiveness Study (NSES) [32] highlighted another instance of bimodality, this time in numeracy scores. According to the NSES findings, one mode of learners (from historically white only schools) demonstrated a numeracy proficiency level of approximately 80%, while the other mode (of learners from traditionally homeland schools) achieved only around 25%.

Through these examples, Spaull demonstrated the complexity and diversity of learning achievement within the context of the nation's education system. Bimodality reflects the significant disparities in educational outcomes between these two groups of learners. It suggests that there are effectively two separate education systems with one serving affluent learners with ample resources, and the other serving disadvantaged learners with limited resources. Recognising this divide is crucial for designing targeted interventions to address the specific needs of learners in each group. Failing to acknowledge bimodality may lead to misunderstandings of educational data and ineffective policy decisions. Therefore, it is essential to adopt strategies that promote educational equity and improve outcomes for all learners.

Recently, an in-depth analysis of the outcomes from the class of 2023 lays bare the profound education crisis confronting South Africa [37]. In 2023, a total of 715 719 full-time learners and 182 056 part-time learners, equating to 897 775 candidates, participated in the NSC examinations across the country. The national pass rate stood at 82.9%, marking a rise from 80.1% in 2022. To pass the NSC examination, learners must attain a minimum achievement mark of 40% in three subjects (including an official home language) and 30% in three other subjects. However, only approximately 40% of all Grade 1 learners eventually progress to sit for the NSC examination in Grade 12 [92]. Among full-time learners, a mere 41 273 individuals achieved scores exceeding 60% for Mathematics. Similarly, a modest 35 468 learners attained marks of at least 60% for Physical Sciences. These figures suggest that less than 6% of NSC examinees demonstrate proficiency in fundamental mathematics or physical science concepts. Consequently, the South African education system appears to be in a dire state, with implications extending to the nation's future economic prospects, as learners exit the educational sphere without adequate preparation. This dilemma poses a significant challenge warranting urgent attention.

1.2 Problem description

The current condition of the South African education system signals a profound crisis, marked by diminishing academic achievements, enduring inequalities among learners, and significant challenges in establishing educational quality and fairness. Assessments such as PIRLS, TIMSS, and reports from research entities like ReSEP and the Reading Panel 2030 make clear the pressing necessity for extensive reforms aimed at addressing enrolment disparities, gender gaps, socio-economic discrepancies, and the fundamental shortcomings in reading, writing, and mathematical competencies among learners. The stark reality of bimodality in academic performance further emphasises systemic inequalities within the educational framework, where one cohort of learners excels while another faces persistent obstacles, perpetuating a cycle of disparity and underperformance. Prompt and decisive action is essential to confront these issues and instigate meaningful transformations in South Africa's educational landscape. Without determined interventions and targeted strategies to enhance educational outcomes and champion equity, the nation risks significant repercussions for its future economic advancement and societal progress. Consequently, urgent systemic reforms and sustained endeavours are imperative to ensure that all South African children have access to high-quality education and the opportunity to thrive academically, thereby securing the nation's future prosperity and societal well-being.

These challenges have prompted an in-depth analysis to explore their underlying complexity. The Systems Thinking for Education Policy (STEP) [95] research group, based in the Department of Logistics at Stellenbosch University, is actively engaged in this endeavour. Their focus revolves around the escalating decline in basic education and the widening disparities in academic opportunities within South Africa. Using sophisticated simulation modelling techniques (such as system dynamics, agent-based modelling, and statistical analysis) STEP researchers aim to conduct a thorough examination of the current education system and propose solutions to enhance its effectiveness.

In 2020, Venter pioneered the development of the Primary School Model (PSM), tailored to simulate public primary schools in the Western Cape province [99]. This model focused on discerning the determinants of academic performance from Grade 3 to Grade 6.

Expanding on Venter's groundwork, Slamang further refined the PSM into the High School Model (HSM) during the same year [83]. This expanded model incorporated additional factors influencing academic outcomes across Grades 6, 9, and 12. Slamang identified ten pivotal factors from literature. These included the quality of school resources, class size, teacher effectiveness, engagement time, economic status, familial and community health, learner motivation, learner attrition rate, social support, and the preparedness of learners for high school at the time of enrolment. Slamang attempted to simulate not only the physical progression of learners through successive grades upon completion of required assessments, but also the evolution of their cognitive capacities, measured through literacy and numeracy metrics.

In 2024, Becker contributed to the refinement and validation of the HSM mapping [12]. Through rigorous statistical analysis and qualitative methodologies, Becker enhanced confidence in the accuracy of the relational mapping process.

Learners and schools are typically categorised into quintiles based on SES, aligning such that quintiles of learners and schools correspond. This alignment, while not legally mandated, often occurs due to factors such as zoning laws and residential proximity, resulting in a homogenous system where learners share similar SES backgrounds within their respective schools and peer groups. Scenarios arise, however, where learners from lower quintiles attend higher quintile schools, creating a heterogeneous system. For example, learners from lower SES backgrounds may gain access to higher quintile schools by means of bursaries or donations. Alternatively, parents might adopt a self-sacrificial rearing strategy, foregoing financial and emotional security at home in an effort to afford an education beyond their means. Such scenarios introduce greater diversity into a cohort that must be served by a single, predetermined system, often designed to accommodate the (hypothetical) average learner. This situation prompts exploration of the emotional, academic, and physical adjustment of lower quintile learners within higher quintile environments.

Departing from the assumption that broad policies targeting the average learner suffice, this study adopts advanced statistical techniques to model how the overarching system accommodates heterogeneous individuals. Although this study does not include the explicit modelling of cognitive development (in the form of literacy and numeracy proficiency), the assumption is made that a physically progressing learner has achieved at least the minimum development required to pass the assessments necessary for grade progression.

Embedded within this research is an acknowledgment of the system's bimodal nature, as discussed by Spaull [92]. The study seeks to explore how learners from disadvantaged backgrounds thrive within an educational environment designed for the nonexistent average learner. Conversely, it examines the potential frustrations experienced by high-performing learners. Through this comprehensive approach, the research aims to provide insights into the efficacy of educational policies and interventions in addressing the diverse needs of learners within the South African context. The assumption that the national government could resolve South Africa's education crisis solely by increasing funding and sending all learners to quintile four and five schools overlooks the intricate factors influencing academic success. Thus, a comprehensive approach must recognise these multifaceted dynamics, emphasising the creation of inclusive learning environments conducive to academic achievement.

1.3 Research questions

This study seeks to address the following research questions:

1. Which latent factors characterise public high school learners from lower socio-economic backgrounds across two contrasting schooling contexts in South Africa?
2. How are these factors interrelated within and between these contexts, and what systemic patterns emerge from these relationships?
3. To what extent do these factors predict academic progression within the broader public education system?
4. Building on the findings of the preceding exploratory questions, how do the characteristics, interrelationships, and predictive significance of these factors manifest within a case study school serving lower socio-economic learners, and how do these patterns compare to those observed at the system level?

1.4 Aim and objectives

The aim of this thesis is to develop a systems-informed understanding of the factors influencing learner progression within the South African education system, particularly among learners from lower socio-economic backgrounds. To achieve this, the following objectives are pursued:

I Analyse literature pertaining to:

   (a) existing modelling methodologies that adopt a systems approach within educational contexts,

   (b) a justification of the methodological approach, and

   (c) simulation methodologies applied in the context of the South African education system.

II To map the factors influencing the South African public high school system, with a specific focus on learners from lower socio-economic backgrounds, by:

   (a) conducting data selection by creating two distinct datasets (one comprising lower socio-economic learners attending lower socio-economic schools, and the other comprising lower socio-economic learners attending higher socio-economic schools) and performing data wrangling on these datasets,

   (b) applying factor analysis to identify latent constructs within the datasets,

   (c) conducting multivariate regression analysis to identify relationships between a dependent variable and multiple predictors, and
   (d) employing structural equation modelling to explore complex interrelationships between variables by conceptualising the system holistically.

III Design a survey comprising the most pertinent questions to ask lower socio-economic learners in South Africa, with the objective of predicting their school progression by:
   (a) applying machine learning techniques to rank the factors by their relative importance, and
   (b) selecting the most appropriate questions based on the outcomes of the machine learning analysis, consultations with the school, and insights from existing literature.

IV To map the factors influencing the learners within the case study school by:
   (a) applying factor analysis to identify latent constructs within the dataset,
   (b) conducting multivariate regression analysis to identify relationships between a dependent variable and multiple predictors,
   (c) employing structural equation modelling to explore complex interrelationships between variables by conceptualising the system holistically, and
   (d) applying machine learning techniques to rank the factors by their relative importance.

V Critically evaluate the findings from the analyses of the public high school system and the case study school in order to inform the development of a decision-support hierarchical framework intended to guide educational interventions for learners from lower socio-economic backgrounds in South Africa.

VI To consolidate the findings of the study, critically reflect on its limitations, and propose avenues for future research within this domain.

1.5 Research methodology

In the domain of complex systems modelling, a range of analytical techniques is available to support comprehensive investigation. This study adopts a quantitative, explanatory case study design, drawing on secondary survey data and advanced statistical modelling techniques to examine the determinants of learner progression among lower socio-economic learners. The integration of these methods enables a nuanced understanding of the interactions between key variables, offering insight into their relationships within the broader systemic context under review.

1.5.1 Methodological approach

The methodological approach adopted in this study comprises a sequence of interrelated analytical techniques, collectively referred to as variate analysis, each selected to address a specific aspect of the research objectives within the broader systems framework. This approach is grounded in the methodological foundation outlined by O'Rourke and Atcher [67], and has since been applied and refined in the context of education systems research by Venter [100], Becker [12], and van der Heever *et al.* [97] across three distinct studies. Their work demonstrated the relevance and efficacy of this analytical structure in addressing the multifaceted nature of educational data, thereby justifying its adoption in the present study:

1. Factor analysis (FA) is employed to extract latent constructs from the General Household Survey (GHS) and the case study dataset.
2. Multivariate regression analysis (MRA) is used to evaluate the relationships between latent constructs and learner progression outcomes.
3. Structural equation modelling (SEM) is applied to model and visualise the complex interdependencies among the identified factors in a holistic manner.
4. Machine learning (ML) is used to conduct a factor importance analysis to identify the most influential variables predicting learner progression.
5. Case study application: The independent school under review is analysed using the same methodological framework to assess whether system-level patterns hold at the school level.

1.5.2 Study data

The primary data source for this research is the South African GHS, publicly accessible via the DataFirst portals, covering the period from 2010 to 2024 [33]. Conducted annually since 2002 by Statistics South Africa, the GHS offers a comprehensive overview of household dynamics, including education, health, and housing. Data are collected through in-person interviews. The survey employs a master sample frame based on the 2011 national census, ensuring representation across both provincial and urban–rural divides.

Given the methodological disruptions caused by the COVID-19 pandemic, the 2020 and 2021 datasets are excluded from this study due to their deviation from typical conditions and that at the outset of the study, the 2022 dataset was the most recent version available. Consequently, the 2022 dataset is deemed the most appropriate for analysis. Data pertaining to the case study school were provided directly by the institution, and as such, no primary data collection was required for this component of the study.

Figure 1.1 provides a visual overview of the methodological approach adopted in this study, illustrating how each method corresponds to the research objectives and contributes to the overall analytical framework.

1.6 Limitations and delimitations

This study is constrained by the data sources available for analysis. Both the publicly accessible GHS and the case study school rely on survey-based data collection methods, which introduce the potential for subjectivity in respondents' answers.

The delimitations of the study include the target population, geographical scope, data sources, and chosen methodology. A deliberate decision was made to focus exclusively on learners from lower socio-economic backgrounds in South Africa, specifically those attending lower quintile public high schools or the selected case study school. These parameters were intentionally selected to align with and address the research questions guiding the study. Geographically, learners from across the country were included, as limited available data did not allow regional segmentation. The 2022 GHS data set was selected as it represented the most recent data available at the start of the study and was considered the first dataset collected under normal conditions following the disruptions caused by the COVID-19 pandemic. The methodological approach was informed and shaped by existing literature and the research aims of the study.

1.7 Ethical considerations

In the pursuit of assessing educational system improvements, the ethical integrity of using socio-economic and household data is of paramount importance. The author affirms a steadfast commitment to ethical principles, ensuring that the research adheres to the highest standards of integrity and respect for individuals' rights. By emphasising that this data is freely accessible to any analyst, the author reinforces the transparency and inclusivity of the research process. It is crucial to note that data gathering is overseen by designated government officials. This arrangement shows a commitment to accountability and integrity in the data collection process. Moreover, stringent measures are taken to safeguard privacy and confidentiality, reassuring stakeholders that the data extracted from sources, (*i.e.*, GHS) contains no personally identifiable information regarding South African learners. The meticulous depersonalisation of data within these sources serves as a testament to the author's unwavering dedication to ethical practices. Through sensitive analysis, efforts are made to prevent the unfair labeling of certain learner groups based on socio-economic status or household circumstances. Above all, the ethical use of data remains grounded in the constructive purpose of identifying areas of educational need and designing interventions for positive impact. Through responsible data analysis, the author ensures that the research serves the betterment of educational outcomes while upholding the highest standards of integrity and respect for individuals' rights. This ethical approach shows the author's dedication to conducting research that not only produces valuable insights but also upholds ethical principles at every stage of the process.

1.8 Publication disclaimer

Chapters 3 and 4 have been compiled as a research article entitled *Feature analysis of learner adaptation across the socio-economic divide in South African public high schools*, which has been reviewed and accepted for publication in the South African Journal of Industrial Engineering. Publication information is currently pending.

1.9 Chapter 1 summary

This chapter introduced the broader context and motivation for the study, focusing on the systemic challenges facing the South African public high school system. It outlined the educational inequalities rooted in socio-economic disparities and the persistent bimodality observed in learner performance. Drawing on national and international assessments, as well as recent reports from academic and policy institutions, the chapter illustrated the need for nuanced and data-driven approaches to improve educational outcomes.

The research questions guiding this study were presented, focusing on the identification, interrelation, and predictive value of key factors influencing learner progression within heterogeneous schooling contexts. The scope and objectives of the study were then detailed, including the application of statistical modelling and machine learning techniques to analyse learner data.

These elements establish a strong foundation for the subsequent chapters, which review existing literature, describe the methodology, present the analysis, and extract the implications of the study.

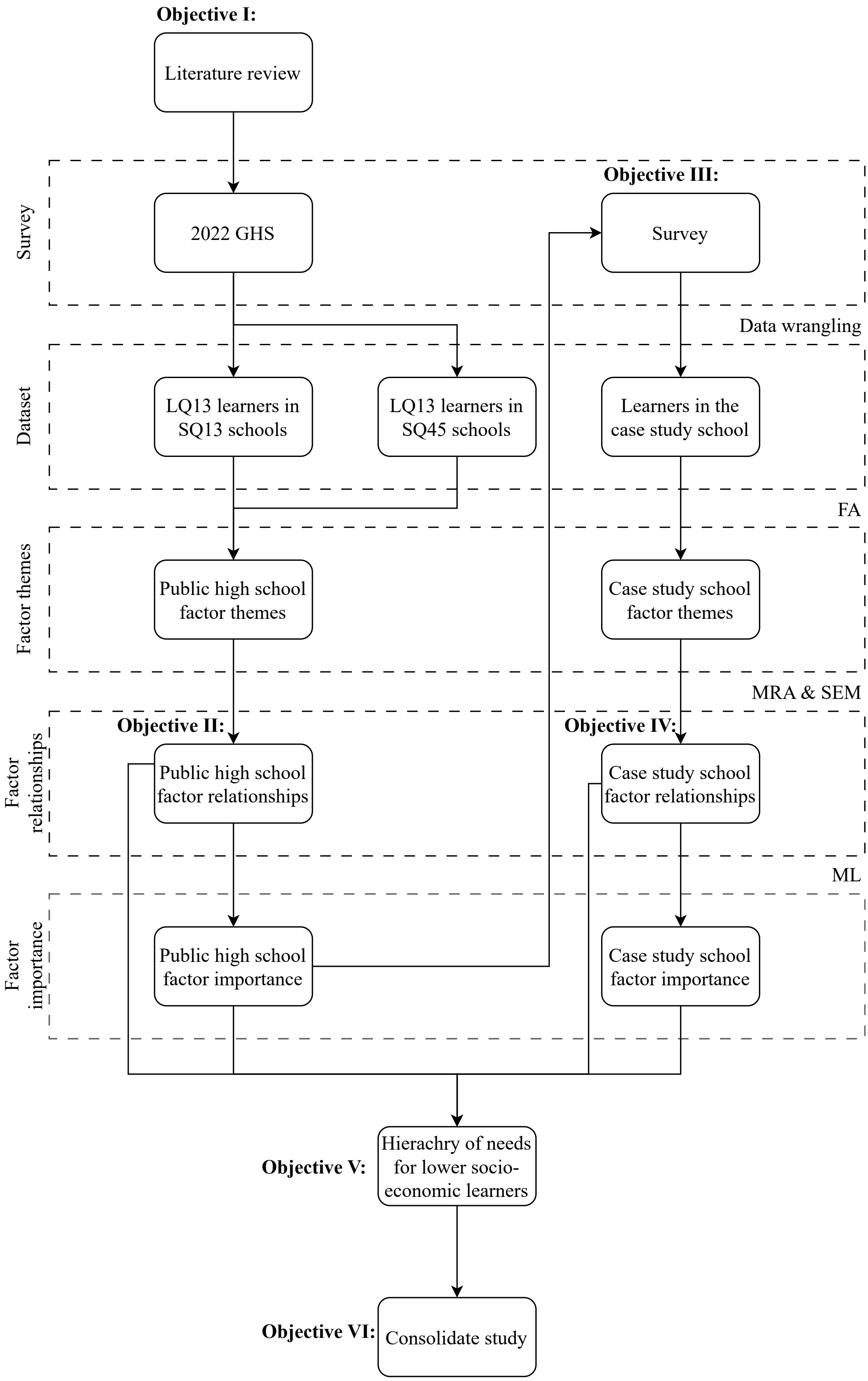

FIGURE 1.1: *Mapping of methodological components to research objectives.*

CHAPTER 2

Contextualisation of the study within systems-based approaches to educational modelling

Contents

This literature review contextualises the study within systems-based approaches to educational modelling in pursuit of Objective I. The chapter begins by examining international case studies that employ a systems approach to understand and improve education systems in Section 2.1. In Section 2.2, the methodological framework employed is substantiated through a critical review of alternative approaches. Finally, the present study is contextualised within the body of work produced by the STEP research group, outlining prior contributions and methodological developments in Section 2.3. This provides a foundation for the current research, situating it within a broader academic trajectory and showing its relevance to ongoing efforts to model and enhance learner progression in the South African education context.

2.1 Modelling education systems using a systems approach

Various researchers have sought to map education systems across diverse international contexts to comprehend their underlying structural complexities. Using system dynamics apparoaches, particularly through the application of causal loop diagrammes (CLDs), these studies reveal feedback mechanisms, interdependencies, and unintended consequences. This approach offers valuable insights into policy formulation, reform implementation, and systemic resilience. The following section briefly reviews a selection of recent papers that illustrate how researchers have applied system dynamics within this domain.

In 2024, Liew *et al.* applied a system dynamics perspective to analyse the integration of artificial intelligence (AI) into higher education systems [56]. Using CLDs, the authors mapped the feedback loops between key components such as technological advancement, pedagogy, institutional structures, and societal expectations. The CLDs were constructed from empirical data and existing literature to reflect real-world complexities. By modelling these interdependencies, the study demonstrated how AI influences student learning, faculty responsibilities, and administrative practices. The systems approach enables a nuanced understanding of the structural and behavioural shifts that accompany AI adoption in higher education.

In 2025, Ramasu and Kanakana-Katumba employed a system dynamics approach to examine the complex dynamics of implementing fee-free higher education in South Africa [70]. Through literature review, stakeholder interviews, and participatory model testing, the authors constructed CLDs that integrate both scientific and experiential knowledge. The diagrammes illustrated key feedback loops and structural interdependencies within the policy environment. By the prioritisation of qualitative analysis and conceptual mapping, this systems approach effectively captures the multifaceted nature of education policy, offering valuable insights into the potential consequences and trade-offs associated with fee-free higher education.

In 2022, Constanza adopted a qualitative system dynamics approach to analyse the local-level impacts of COVID-19 pandemic-related educational policies in Italy, focusing on state schools in Palermo [27]. Using CLDs, the researcher illustrated the interactions between policy decisions, socio-economic factors, managerial discretion, and the role of street-level bureaucrats. The CLDs were developed through a multi-step process involving literature review, stakeholder engagement, iterative model building, and validation using SD software. By identifying 31 interacting feedback loops, the study provided a comprehensive systems view of educational service adaptation during the pandemic, offering insights into co-creative processes and policy implementation.

In 2024, Keith *et al.* applied a qualitative system dynamics approach to anticipate the unintended consequences of educational reform in an urban U.S. school district [53]. Through interviews and workshops with educators, parents, and administrators, the authors constructed CLDs to model feedback relationships contributing to declining enrolment and academic outcomes. The CLDs, grounded in fieldwork and literature, reveal a “capability trap” wherein reform efforts are undermined by systemic constraints, leading to policy resistance. This systems-based conceptual model illustrated the complex interplay of trust, resources, and institutional capacity, offering a structured framework for anticipating reform side effects.

In 2021, Ammara *et al.* employed systems thinking to model the complex dynamics of student learning during and beyond the COVID-19 pandemic. Using CLDs and stock and flow diagrammes, the authors explored how systemic structures contribute to persistent educational challenges [4]. The CLDs were developed by identifying key variables from literature and conceptual insights, mapping their cause-and-effect relationships with defined polarities to reveal reinforcing and balancing feedback loops. These diagrammes illustrated interconnections between student motivation, curriculum reforms, and online distractions. Field insights revealed that quick-fix, top-down policies often produce unintended consequences, such as teacher demotivation and reduced academic performance.

The use of CLDs in the reviewed studies demonstrated both the relevance and practical applicability of system dynamics in educational research. However, these studies predominantly relied on existing literature, stakeholder assumptions, or expert input to construct their models. In contrast, the present study aims to develop a systemic view of the South African education system using raw empirical data as its foundation. Consequently, the methodological approaches employed in the aforementioned studies are not directly transferable to this context.

2.2 Methodological justification

The methodological framework employed in this study is based on the approach outlined by O'Rourke and Hatcher [67]. Their work presents a comprehensive methodology for progressing from raw data to the extraction of latent constructs using exploratory and confirmatory FA, followed by the identification of significant relationships among these constructs through MRA and SEM. Prior to adopting O'Rourke and Hatcher's framework, the techniques they propose were critically assessed and compared with alternative methods capable of yielding comparable analytical outcomes.

2.2.1 Alternative approaches for factor theme identification

To identify latent factor themes in the data, O'Rourke and Hatcher used FA which is a well established method. Alternative methods which can be used are thematic analysis, principle component analysis (PCA), or latent class analysis (LCA).

Thematic analysis comprises a set of methodological approaches used to identify, analyse, and interpret patterns of meaning within qualitative data [17]. The data employed in this research project is quantitative in nature; therefore, thematic analysis is not an appropriate methodological approach.

PCA is a statistical method employed to reduce the dimensionality of large datasets by transforming the original variables into a smaller set of principal components that capture the greatest amount of variance [41]. Although PCA is effective for simplifying data, it does not reveal the underlying structure or conceptual relationships between variables [78]. In contrast, exploratory FA is designed to identify latent constructs that account for observed correlations. Given that this study seeks to explore the underlying themes shaping the education system, PCA is deemed unsuitable, as it focuses solely on variance reduction rather than uncovering deeper theoretical structures within the data.

LCA is a statistical method used to identify unobserved subgroups within a population based on individuals' responses to categorical items, typically in the form of binary options such as "yes" or "no" [82]. These latent groups are not immediately apparent but are inferred from patterns in the data. LCA enables researchers to classify individuals who exhibit similar response patterns, thereby facilitating a more nuanced understanding of heterogeneity within the population. In contrast to FA, which seeks to identify continuous underlying dimensions that explain correlations among items, LCA is concerned with categorising individuals into distinct, non-overlapping classes based on their observed responses. Although LCA could offer valuable insights into the types of learners present, this research is primarily concerned with the broader system in which these learners exist and the nature of their interactions within that system.

2.2.2 Alternative approaches for factor relationship identification

To identify latent factor relationships within the data, O'Rourke and Hatcher employed MRA and SEM to represent the relationships among the latent factors. An alternative approach that may be used for this purpose is network analysis.

Network analysis is a statistical method used to explore and visualise the relationships between multiple variables by representing them as a network of variables and connections between variables [16]. Rather than reducing data or identifying hidden traits, this approach focuses on understanding how variables directly relate to one another. Network analysis is particularly useful for identifying which variables are most central or influential, evaluating how robust these relationships are, and gaining insight into the overall structure of the data. It can be applied to both single-time-point data and data collected over time, offering flexibility in how relationships are examined.

In this case, both the MRA and SEM approach proposed by O'Rourke and Hatcher and network analysis appear suitable for application in the current study. However, given the intention to follow and validate O'Rourke and Hatcher methodological framework, this study proceeds with their approach.

2.3 Foundations in prior modelling of South African education

The South African education system grapples with substantial challenges, necessitating in-depth investigation to uncover its underlying complexities. Engaged in this effort is the STEP research group, situated in the Department of Logistics at Stellenbosch University. The group's primary focus centres on addressing the escalating deterioration in basic education and the widening disparities in academic opportunity across South Africa. Employing advanced simulation modelling techniques, most notably system dynamics (SD), agent-based modelling (ABM), and statistical analysis—STEP researchers aim to comprehensively examine the education landscape and propose data-informed solutions to improve its efficacy.

Venter [99] explored the intricate dynamics of the South African basic education system, employing a systemic approach to understand the causal linkages and interventions influencing learner performance. Through the development of simulation models, Venter investigated key factors related to school leadership, teacher effectiveness, early childhood development, and primary school achievement. The outcomes of this work illuminated the complexity of the system and showed the need for holistic strategies to improve educational outcomes.

The primary aims of Venter's dissertation encompassed several facets of the education system in South Africa. Firstly, the influence of school leadership on academic performance was analysed to better understand systemic dynamics. Secondly, the study examined the impact of teacher quantity and quality, with a focus on interventions to improve effectiveness. It also addressed early childhood development, simulating developmental trajectories and identifying factors affecting school readiness. Finally, factors influencing academic achievement in Grade 3 and Grade 6 were analysed, culminating in the development of the Primary School Model (PSM). The CLD for the PSM is shown in Figure 2.1.

Venter advocated for the adoption of systems thinking, emphasising the recognition of interconnected factors influencing learner outcomes as an imperative strategy. Additionally, Venter emphasised the importance of implementing equitable initiatives, including measures to alleviate community poverty and the strategic allocation of resources to enhance both classroom

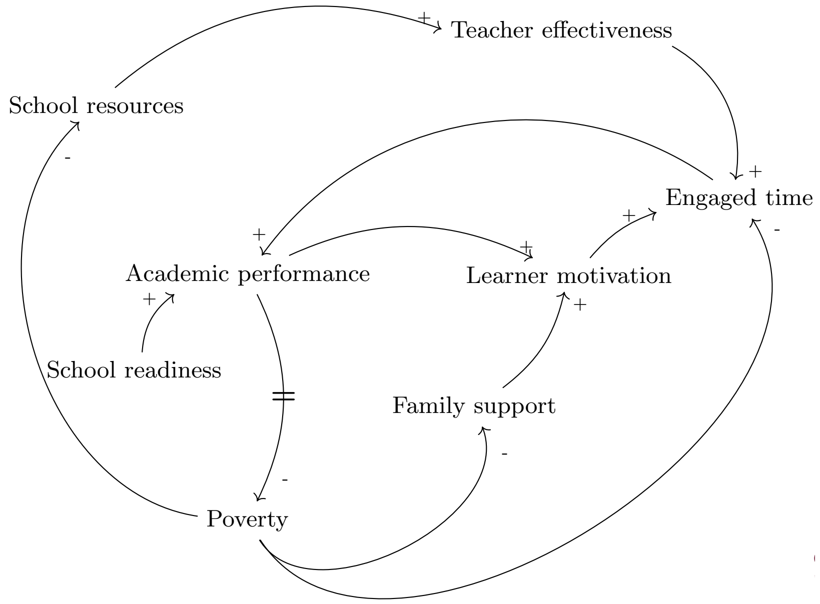

Figure 2.1: *The CLD for learner achievement at primary school level as per Venter[99].*

and home environments. Notably, the overarching finding of Venter's research suggests that, in South African public primary schools, the most substantial influence on academic achievement occurs when there is an improvement in home support and structure. Primary school learners are significantly influenced by their home environments, showing the critical role of family support in educational success.

Expanding on Venter's groundwork, Slamang [83] provided a comprehensive analysis of the factors influencing academic performance in public high schools in the Western Cape region of South Africa. The study employed a system dynamics approach to simulate the interactions of various determinants over time within the public high school system.

Slamang identified key factors affecting academic performance, particularly in Grades 9 and 12, which are pivotal for progression to higher education. These factors include school resources, class size, teacher effectiveness, engagement time, economic status, familial and community health, learner motivation, learner attrition rate, social support, and the preparedness of learners for high school at the time of enrolment. A visual depiction of the CLD is presented in Figure 2.2. The data for these factors were collected from the GHS, which provided insights into the perceptions of learners attending public schools in the Western Cape.

The study explored interventions to enhance academic performance and address disparities in the Western Cape high school system, focusing on home-based, classroom, and poverty-related strategies. Home-based interventions aimed to supplement the educational experience by enhancing parental involvement, providing additional learning resources, and addressing socio-economic challenges. Classroom interventions targeted the school environment, striving to improve teacher effectiveness, optimise resources, and foster a conducive learning atmosphere. Crucially, the study emphasised the need to address poverty-related challenges, advocating for

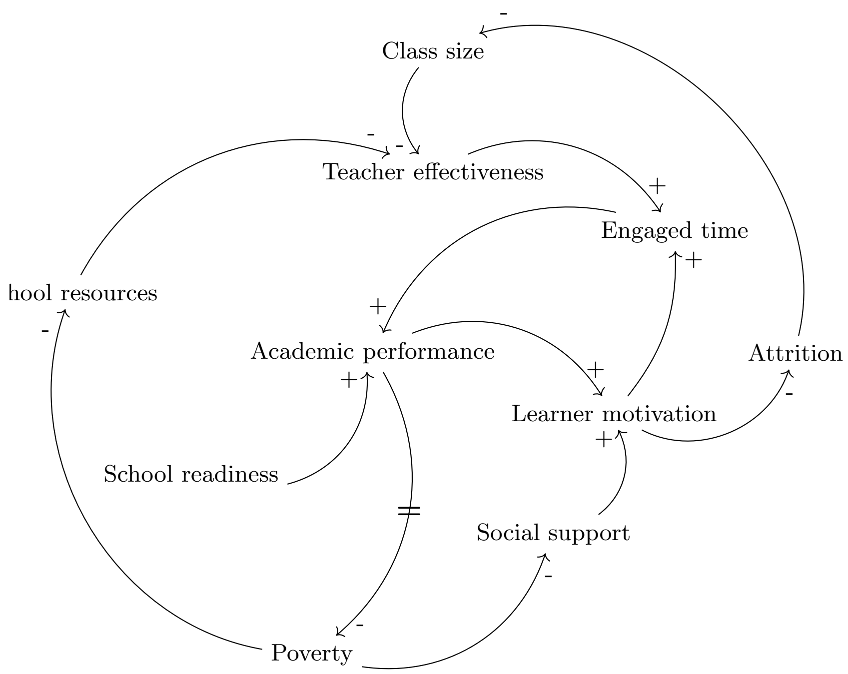

FIGURE 2.2: *The CLD for high school learner achievement. Determinants and causality for learner achievement in the high school system as per Slamang [83].*

socio-economic support, access to basic needs, and equal opportunities for all learners, thereby fostering a more equitable learning environment.

Slamang noted that, in the Western Cape high school system, the intervention that had the most profound effect on academic performance was the classroom intervention. This involved increasing the number of teachers and enhancing both school resources and teacher effectiveness. By improving elements such as teacher salaries, the quality of school facilities, and the provision of learning materials, this intervention significantly improved the standard of education and the level of student engagement in the classroom. A notable divergence from Venter's research is the emphasis placed on the influence of classroom interventions, as opposed to interventions at home or those addressing poverty. While Venter emphasised the significance of home-based interventions in boosting academic performance at the primary school level, Slamang showed the crucial role of classroom interventions at the secondary school level.

Subsequent to Slamang, in 2024, Becker conducted a comprehensive investigation into the South African public high school system, with a particular focus on identifying and understanding the factors influencing learner progression from Grade 8 through to Grade 12 [12]. The principal objective of this study was to refine and validate a causal mapping of these contributing factors within the broader context of the South African education landscape.

Drawing on statistical analysis and using data from the 2019 GHS, Becker identified nine critical factors that significantly affect learner progression. These include supported learner retention, academic performance, school resource availability, family structure, household extravagance, food insecurity, household wealth, access to basic utilities, and environmental challenges. These factors are presented as a relational mapping in Figure 2.3.

The findings reveal that learners from stable family structures and wealthier households are more likely to experience improved academic outcomes. Furthermore, reduced levels of food insecurity

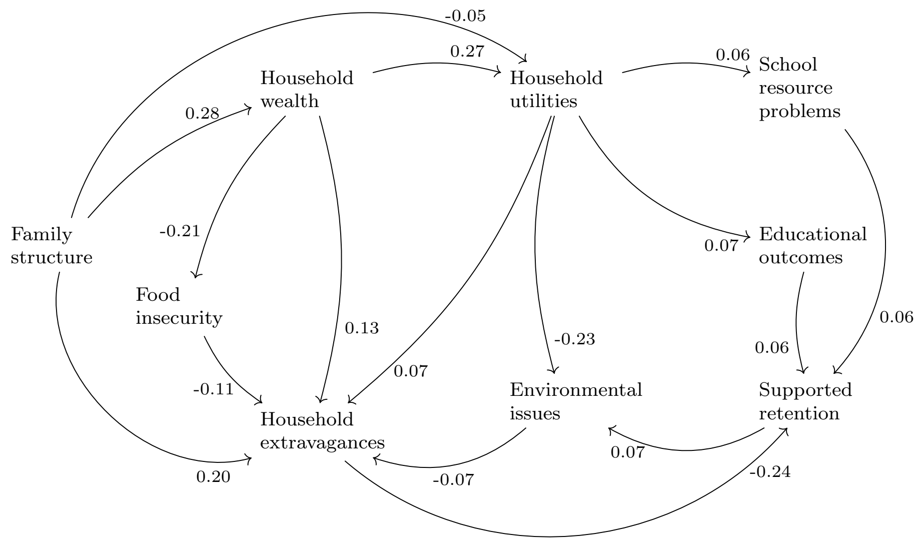

Figure 2.3: *Relational mapping of the significant factors via path modelling of the 2019 GHS as per Becker [12].*

and consistent access to utilities are positively associated with learner retention. These results show the profound impact of socio-economic stability on educational success.

Although health-related variables did not emerge as prominent direct determinants, environmental factors, most notably food insecurity and household wealth, proved to be stronger predictors of learner outcomes than previously assumed. This points to a growing recognition of the household and community context as critical determinants in shaping learners' educational trajectories.

While both Slamang and Becker have contributed valuable insights into the systemic challenges and enablers within the South African education system, it is also imperative to consider the influential work of Spaull [92]. Spaull delineates a marked bimodality within the system, characterised by a pronounced division of learners into two distinct educational cohorts. This bifurcation is largely based on socio-economic status, language of instruction, and school typology.

In light of the combined findings of Slamang, Becker, and Spaull, there is a compelling rationale for exploring the systemic underpinnings of this dual education reality. By employing cohesive methodologies, particularly those informed by the approaches of Slamang and Becker, deeper insights into the structural dynamics sustaining these disparate schooling systems may be gained. Such an analytical perspective could prove instrumental in formulating holistic strategies aimed at addressing inequality and fostering inclusive educational development in South Africa.

2.4 Chapter 2 summary

This literature review contextualised the study within systems-based approaches to educational modelling in pursuit of Objective I. The chapter began by examining international case studies that employed a systems approach to understand and improve education systems. Thereafter, the methodological framework used in the present study was substantiated through a critical review of alternative approaches. Finally, the study was positioned within the body of work produced by the STEP research group, outlining prior contributions and methodological developments. This provided a foundation for the current research, situating it within a broader academic trajectory and demonstrating its relevance to ongoing efforts to model and enhance learner progression in the South African education context.

CHAPTER 3

Variate analysis of the South African public high school system

Contents

In this chapter, variate relationships are investigated within the context of the South African education system as a means to achieve Objective II. Section 3.1 introduces the GHS, followed by a preliminary examination of the dataset to identify patterns, trends, and potential relationships. Subsequently, Section 3.3 provides a detailed account of the data selection and wrangling processes employed in this study.

In Section 3.4, the theoretical framework underpinning the methodologies used to determine variate relationships is outlined. These methodologies encompass FA, MRA, and SEM. The application of these techniques to two distinct datasets is presented in Sections 3.5 and 3.6. In Section 3.7, the learner profile is unpacked to deepen the understanding of the characteristics and circumstances of these learners.

3.1 Introduction to the General Household Survey

The findings in this study are based on an analysis of the 2022 edition of the GHS. This annual survey, conducted through face-to-face interviews, collects data on a broad spectrum of topics related to households and individuals in South Africa. Household-level characteristics include household type, home ownership, access to water and sanitation, service access, transportation, household assets, land ownership, and agricultural production. Individual-level characteristics encompass demographic details, the respondent's relationship to the household head, marital status, language proficiency, education level, employment status, income, health, fertility, mortality, disability, and access to social services [33].

The GHS has been conducted annually since 2002 by Statistics South Africa (StatsSA) and employs a master sample frame to ensure representativeness at both provincial and urban-rural levels. The dataset is publicly accessible via the DataFirst open data portal hosted by the University of Cape Town [34]. With an average response rate of 88.6%, approximately 20 000 households are surveyed each year. The GHS data were processed using the Statistical Analysis System (SAS) software, selected for its comprehensive capabilities in data manipulation, analysis, and visualisation [77].

GHS datasets from the COVID-19 period are considered anomalies due to the exceptional circumstances of the pandemic. As such, the 2020 and 2021 datasets are excluded from this study, as numerous variables were likely distorted by lockdown conditions. The 2022 dataset is presumed to be the first to reflect a return to more typical conditions, making it the preferred dataset for this analysis.

3.2 Initial insights of the 2022 GHS

The primary objective of exploratory data analysis is to identify underlying patterns, trends, and anomalies within the data, providing a foundational understanding of the study population. This initial analysis serves as a crucial precursor to subsequent, more in-depth statistical procedures. The descriptive findings presented here are based on the statistical release published by StatsSA alongside the GHS dataset [34].

Household composition in South Africa is highly diverse. Nuclear family units, comprising parents and children, account for 40.1% of households, while extended family arrangements, including two- and three-generation households, constitute 43.8%. Single-person households represent 25.1%, and 4.4% are skip-generation households. These varied structures significantly shape the domestic environments in which learners are raised.

Educational participation rates among youth remain relatively high. While 97.0% of fifteen-year-olds (the terminal age for compulsory schooling) are enrolled in educational institutions, this rate drops to 66.3% for eighteen-year-olds. Notably, 23.4% of twenty-year-olds remain enrolled in high school, reflecting extended educational trajectories. The implementation of no-fee schools has substantially expanded access: in 2022, 67.7% of learners attended such institutions, up from 21.4% in 2007. However, challenges persist. The main reasons cited for early school leaving are academic underperformance (24.2%) and financial constraints (22.4%). Family responsibilities disproportionately affect female learners, with 12.1% leaving school for this reason, compared to just 0.2% of males.

Educational attainment has improved, with the proportion of individuals aged 20 and older who completed Grade 12 rising from 30.5% in 2002 to 50.5% in 2022—though these gains have not always been matched by improvements in educational quality.

Access to basic services and overall living conditions remain critical determinants of educational outcomes. While 89.6% of households were connected to the main electricity supply in 2022, disparities persist in energy access for cooking and heating, particularly in rural areas. Social grants continue to play a vital role, with 37.0% of households receiving support. Housing remains a concern, with 12.3% of the population residing in informal settlements. Environmental challenges (such as declining waste removal services and low recycling rates) also affect overall household wellbeing.

These findings help explain the living conditions of the learners in this study and set the background for analysing their progression within the South African education system.

3.3 Data selection and wrangling

Data wrangling entails the cleaning and transformation of data from the GHS to enable statistical analysis. SAS is used to analyse data from two GHS files, namely `household.csv` and `person.csv`. The household-level file contains data on various aspects of household-level characteristics, including demographics, income, housing conditions, and access to services, while the person-level file contains individual-level data on personal demographics, education, employment, health, and other personal attributes. By merging these files based on unique identifiers, one can link individual characteristics to their household's SES. This integration enables more nuanced statistical analyses by providing the ability to explore the interactions between person-level characteristics and household conditions. It is then possible to examine how household factors influence individual outcomes, and the merged dataset facilitates multilevel modelling, where both individual and household variables can be included to account for hierarchical data structures, leading to more accurate and insightful conclusions. The data were processed using a series of filtering steps to remove extraneous observations and ensure the remaining data aligned with the research objectives. This data cleaning process was essential for obtaining reliable and accurate results.

After the initial data cleaning process, further filters were applied to isolate the target learner population for analysis. This involved isolating individuals between the ages of thirteen and twenty-one years who attended public high schools within South Africa. Additionally, only learners enrolled in Grades 8 through 12 are considered. To account for attrition, individuals of this same age group who were neither enrolled for formal education nor obtained an NSC were included in the analysis. By implementing these filters, the analysis pertains to a specific and relevant group of learners within the broader educational system.

Next, a total household income was calculated and attributed to each learner. Within the context of this analysis, the household income encompasses the total annual monetary inflows received by all members residing in a single household. This included income sources such as grants, salaries, wages, pensions, and any remittances received by household members. The household income was used to determine the socio-economic quintile category of each learner based on their household's economic standing.

This study employed a stratified approach based on learner quintile (LQ) and the school quintile (SQ) that the learner attended. Learners were categorised into two groups: Learners in the lower three quintiles (LQ13) typically describe learners from the lower-income segments of the population, often characterised by limited access to resources, lower educational attainment, and higher vulnerability to economic hardships. These learners originated from households with a total annual household income less than the specified threshold of R15 600, determined by considering the average annual household income [29]. Learners in the upper two quintiles (LQ45) are typically from the higher-income segments of the population, often characterised by greater access to resources, higher educational attainment, and improved living standards. These learners originated from households with a total household income greater than a specified threshold. For the objectives of this study, only LQ13 learners were retained for analysis.

While not formally mandated, there is a tendency for learner quintile to align with school quintile due to socio-economic factors. Quintile 1 to 3 schools (SQ13) are no-fee schools, receiving the highest level of government funding. Quintile 4 to 5 schools (SQ45) charge fees and serve more affluent communities, receiving lower levels of government funding due to their better access to resources and financial support from the surrounding communities [24]. To reflect this alignment and generate a meaningful union between learner and school quintiles, learners in the merged dataset were each assigned a quintile identifier based on their calculated total household income. This approach aimed to capture the perspectives of learners within their specific school contexts, particularly those from lower socio-economic backgrounds (LQ13) attending either low-resource (SQ13) or high-resource (SQ45) schools, respectively. This process enables the analysis of learner characteristics while considering both their household income and the school's socio-economic context.

To incorporate the influence of attrition, it was assumed that learners who have left the education system due to financial constraints (given that SQ13 schools are typically fee-exempt), academic challenges (as SQ45 schools often have higher academic standards), geographical inaccessibility (since LQ13 learners are more likely to reside near SQ13 schools), or social barriers to entry (as SQ45 schools may have more stringent admission criteria) would have previously attended SQ45 schools. While acknowledging the inherent subjectivity of this assumption, it was considered essential to account for attrition in the analysis where this metric is not immediately captured in the 2022 GHS. This approach was deemed the most appropriate method given the available data and research constraints.

The filtered dataset consisted of 302 features, including categorical data. These categorical variables were transformed into binary indicators through a process of one-hot encoding. This involved the generation of a binary variable for each feature within each categorical variable. For instance, the question "What means of transport is usually used by the learner to get to school?" with nine possible responses was transformed into a binary variable to most appropriately represent the data space. The prompt "Does the learner walk to school?" would be represented by a value of 1 for a "yes" response and a value of 0 for a "no" response. Generated variables, termed indicators, store this binary encoded data. Data processing was applied in this way for all relevant survey prompts, adhering to the survey structure and addressing sections within the GHS. The generation of indicators may involve some subjectivity, as decisions regarding their construction required subject matter expertise and personal judgment. The categorisation for the features from the GHS into relative indicators is displayed in Table A.1 for the person-level indicators and in Table A.2 for the household-level indicators in Appendix A.

The GHS data presents a hierarchical structure, with learners nested within households. To fully capture the complexities of this structure, multilevel modelling is the ideal analytical approach [69]. To address the hierarchical nature of the data while maintaining analytical tractability, a random sampling technique was employed. Within households containing multiple learners, a single learner was randomly selected to represent the household. This approach flattens the data structure, enabling the application of traditional single-level analysis methods. Becker demonstrated that household members, particularly siblings, often share similar experiences and influences [12]. This suggests that analysing a single learner within a household can provide valuable insights into the broader dynamics and challenges faced by the entire family unit. While acknowledging the potential limitations of this simplification, it is believed that this approach provides a reasonable approximation of the underlying data structure for the purposes of this study.

3.4 Theoretical framework of techniques

To establish variate relationships within the GHS dataset, a series of analytical techniques was employed. FA is firstly used to reduce the high dimensionality of the data. By identifying underlying latent factors, FA reduces the complexity of the dataset, making it more manageable for subsequent analyses. MRA is then employed to examine the relationships between the resultant factors to determine the nature of their relationships. Finally, SEM is applied to estimate the strength and direction of variate relationships between the factors, providing a more comprehensive understanding of the underlying processes and thereby offering a potential causal loop diagramme with which the public high school system may be described.

3.4.1 Factor analysis

FA is a method for dimensionality reduction, transforming complex datasets into fewer and more meaningful features [96]. The technique operates by uncovering latent patterns within a set of observed variables. By identifying a smaller set of underlying factors that encapsulate the key information, FA facilitates more efficient data analysis [54]. FA provides a summary of multiple observed variables to capture a broader concept inherent in the data by uncovering underlying structures and themes concealed within large datasets. FA consists of two distinct approaches, namely exploratory FA (EFA) and confirmatory FA (CFA). EFA, originally developed by Charles Spearman [93] in the area of human abilities, serves as a method for uncovering the underlying structure within a dataset, operating without pre-established hypotheses. Originally, FA was merely an exploratory statistical method, but in 1973, Jöreskog [50] developed a method to test and validate specific hypotheses regarding the data's structure and defined this method as CFA.

At the initialisation of the FA process, the suitability of the technique for the specific dataset is examined by means of two distinct tests [101]. Firstly, Bartlett's test of sphericity evaluates the overall strength and significance of correlations within a dataset [11]. This test focuses on the null hypothesis, which posits that the correlation matrix (a representation of all pairwise correlations between variables) is an identity matrix. In the context of the correlation matrix, an identity matrix signifies that all variables are independent, exhibiting no correlation. The test aims to assess whether this assumption holds true. A statistically significant result, typically indicated by a p-value less than 0.05, leads to the rejection of the null hypothesis. This rejection implies that the correlation matrix deviates significantly from an identity matrix, suggesting the presence of meaningful relationships or structure within the data. Such structure is a fundamental requirement for the successful application of EFA.

Secondly, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (MSA) serves as a statistical index that quantifies the proportion of variance within each variable in a dataset that can be explained, without error, by the remaining variables [52]. Essentially, the KMO measure assesses the extent to which variables in the dataset share common underlying factors. A KMO-MSA value closer to 1 indicates a stronger likelihood that FA is an appropriate technique for the data. Conversely, values below 0.5 suggest that the dataset may not be well-suited for FA due to a lack of shared variance among variables. Values ranging from 0.6 to 0.8 suggest that FA might be cautiously considered, with a higher potential for successful analysis as the KMO value approaches 0.8. Values exceeding 0.8 are generally considered favourable for FA, suggesting a high degree of shared variance that can be effectively captured through FA.

Subsequently, the determination of the number of factors to retain in FA involves two distinct methods. A factor is a dimension or construct that is a condensed statement of the relationships between a set of variables [54]. Firstly, the eigenvalue-greater-than-one rule can be applied, where the number of factors to retain corresponds to the number of factors that adhere to this rule [51]. Alternatively, a scree plot can be constructed, with the number of factors on the x-axis and the eigenvalues on the y-axis [25]. A key point in this plot is where the curve's slope changes abruptly, known as the elbow of the curve. The factor number at this point of change indicates the number of factors to retain, as visually depicted in Figure 3.1.

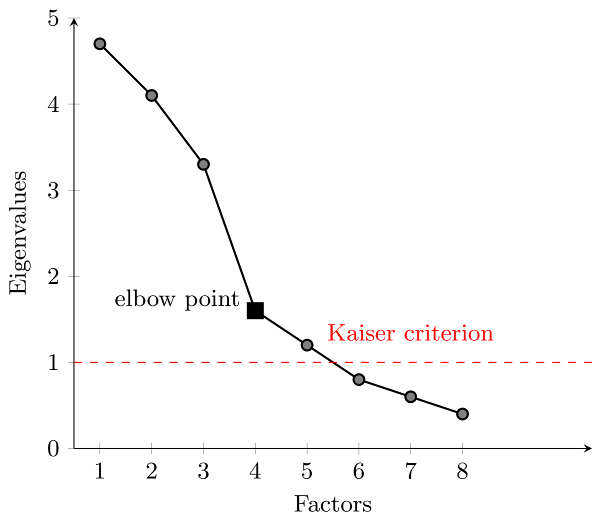

Figure 3.1: *An example of a scree plot to determine the appropriate number of factors to retain during EFA (adapted from [25]).*

EFA is undertaken when suitable variables and a desired number of factors are specified. The EFA output comprises a rotated factor pattern, essential for interpreting the underlying data structure. Rotated factor patterns achieve "simple structure," where each variable loads heavily on one factor and minimally on others, simplifying interpretation [67]. Varimax rotation, an orthogonal technique, produces uncorrelated factors and aims to maximise the differences in factor-variable associations. This results in each variable being strongly associated with only one factor, making the interpretation clearer by distinctly separating the correlations between factors and variables. These factor loadings are considered high if they exceed 0.6 (regardless of sign) and moderately high if they are above 0.3; other loadings can typically be disregarded [54]. In this way, common conceptual meanings among variables that load on the same factor can be identified, aiding in the understanding of underlying constructs. The rotated factor pattern matrix assists in assigning meaningful themes to factors by examining variables with high loadings. In this way, rotated factor patterns are important in EFA for simplifying factor interpretation, identifying latent constructs, and determining factor names. Ultimately, uncovering the latent structure of the data allows informed analytical decisions to be made.

CFA is employed to evaluate the extent to which a theoretical model aligns with empirical data. A well-fitting model effectively accounts for the observed covariances within the data [67]. Assessing model fit is complex, as no single index provides a definitive evaluation [19]. Consequently, multiple fit indices are typically examined.

The comparative fit index (CFI) compares a specified model to a null model, with values greater than 0.94 indicating a satisfactory fit. The standardised root mean square residual (SRMR) represents the average discrepancy between observed and predicted correlations, with values below 0.055 suggesting a good fit [43]. Finally, the root mean square error of approximation (RMSEA) estimates the model's deviation from population parameters [55]. Values less than 0.055 indicate a close approximation [20].

To build upon the exploratory findings of FA, Cronbach's alpha is computed to verify that the extracted factor structures are internally consistent and reliable for further analysis. Introduced by Lee Cronbach in 1951 [30], this widely adopted metric evaluates the internal consistency of a scale by measuring how closely related the indicators within a factor are. Represented as a value between 0 and 1, higher alpha values denote stronger internal consistency and greater reliability in capturing the same underlying construct.

FA is an iterative process, involving repeated refinements to the model. This approach allows for adjustments and modifications based on the interpretability of the factors and the fit of the CFA model. Through iterative refinement, the extracted factors become more meaningful, offering a more useful representation of the underlying latent constructs in the data. This iterative process improves the overall quality and interpretability of the FA results.

3.4.2 Composite factors for further analysis

FA is employed to reduce the dimensionality of the GHS to fewer latent constructs. However, each factor still consists of multiple significant variables. To calculate a single score per factor, a weighted average approach is implemented. Specifically, factor loadings produced by the FA are multiplied by binary values representing the observed responses per record for each indicator. This process yields composite factor scores ranging from 0 to 1, with higher values indicating more favourable conditions. For instance, a score of 1 for a factor describing family structure would represent a household with both parents married and their biological child residing together. A household experiencing food insecurity, characterised by insufficient food intake, food depletion, or hunger among members, would receive a composite factor score of 0 for the food security construct. The mathematical formulation for the generation of these composite factors is presented in Equation 3.1, where ℓ_j represents the factor loading for indicator j derived from EFA, and ι_j is the binary encoding of indicator j, for the n indicators that comprise each composite factor. Therefore, for all factors F_i,

$$F_i = \frac{\sum_{j=1}^{n} \ell_j \iota_j}{\sum_{j=1}^{n} \ell_j} \quad \forall i. \tag{3.1}$$

3.4.3 Multivariate regression analysis

MRA is a versatile statistical technique employed to examine the relationship between a quantitative dependent variable and a set of independent variables [14]. These independent variables can be either quantitative or categorical, and the model can accommodate both linear and nonlinear relationships. MRA enhances the ability to assess the influence of multiple predictors simultaneously while controlling for the effects of other variables.

Towards the exploration of the relationship between a dependent variable, denoted as Y, and a set of k independent variables (*i.e.*, x_1, x_2, x_3, ..., x_k), MRA is employed [15]. Data for these variables are collected from N cases. For example, one might seek to predict household energy access (Y) based on factors such as household employment (x_1), learner health (x_2), and school support (x_3). To estimate the dependent variable, the model multiplies each independent variable score for a given case by its corresponding coefficient, sums these products, and adds a constant term, as illustrated in Equation 3.2. The resulting value is the predicted value of Y for that case, denoted as $\hat{Y}$. The regression coefficient between the observed Y values and the predicted $\hat{Y}$ values is quantified by the multiple regression coefficient, R. If a high degree of regression coefficient exists among predictor variables within a model, this condition is referred to as multicollinearity. Such circumstances can pose challenges for statistical analysis. To address multicollinearity, techniques such as principal component analysis may be employed [31]. The significance of the relationship derived from MRA is determined through a hypothesis test where $p < 0.001$ is strongly significant and $p < 0.05$ is moderately significant.

$$\hat{Y} = B_0 + B_1 x_1 + B_2 x_2 + B_3 x_3 \tag{3.2}$$

3.4.4 Structural equation modelling

Over a decade ago, methods for modelling the structure of relationships among variables using systems of equations began to gain traction among sociologists, as evidenced by the work of Bagozzi [10], Oliver and Bearden [66], and Shimp and Kavas [81]. These methods evolved into what is now recognised as modern statistical techniques. SEM is defined as a sophisticated technology that investigates the complex relationships between latent or observed factors, in this context, at both the individual and household levels [6].

SEM is applied to the relevant dataset to uncover and explore further relationships between variables. Complementing MRA, SEM facilitates the simultaneous modelling of multiple relationships, thereby providing a holistic view of the entire system. SEM extends traditional regression analysis by incorporating latent variables and their interrelationships, enabling the assessment of both direct and indirect effects within a cohesive framework. Furthermore, SEM's global fit indices offer an overall evaluation of how well the theoretical model corresponds with the observed data. Similar to CFA, model fit is assessed using CFI, SRMR, and RMSEA [67]. As with CFA, CFI is deemed satisfactory when values exceed 0.94. SRMR suggests a good fit when values are below 0.055 [43]. Lastly, RMSEA indicates a close approximation when values are less than 0.055 [20].

The standardised path coefficients (β), which are rescaled, estimate the strength and direction of relationships between variables, measured on a common scale with a mean of zero and a standard deviation of one. In SEM, the standard error (SE) represents the degree of variability associated with an estimated path coefficient. A smaller SE indicates higher precision in the estimate, suggesting that the results are less likely to vary substantially across different samples. The t-value, derived as the ratio of the estimated path coefficient to its standard error, as presented in Equation 3.3, serves to evaluate the statistical significance of the relationship between variables. A higher absolute t-value signifies a stronger relationship and a reduced likelihood that the observed effect is attributable to chance.

$$t = \frac{\beta}{SE} \tag{3.3}$$

3.5 Variate relationship analysis for LQ13 learners in SQ13 schools

The first analysis focuses on LQ13 learners attending SQ13 schools, using a sample of 2 839 learners, comprising 52% male and 48% female. Home languages varied, with 28% speaking isiZulu, 17% Sepedi, 15% isiXhosa, and 11% Sesotho. In an effort to determine variate relationships within the South African public high school system for LQ13 learners in SQ13 schools, this dataset is analysed using FA, MRA, and SEM.

3.5.1 Significant factors for LQ13 learners in SQ13 schools

Towards a comprehensive view of the factors influencing learners, FA was conducted on the dataset pertaining to LQ13 learners in SQ13 schools. This approach enabled an in-depth exploration of the factors influencing individual learners and the broader household environment, providing insights into how these factors collectively shape the learner's educational trajectory.

Firstly, the analysis centres on variables concerning the person-level characteristics of the learner, including sections on education, health, and family structure. EFA was conducted on a predefined set of indicators, which are enumerated in Table A.1 of Appendix A. The suitability of the data for EFA was ascertained with an overall MSA of 0.62, indicating sufficient commonality among the variables to justify the use of FA. In Table 3.1, the relative indicators, their corresponding KMO-MSA values, and the factor loadings for the three factors uncovered are presented. The factor loadings are multiplied by a factor of 100 to better display their clustering. CFA results yielded ideal fit indices (CFI = 0.92, SRMR = 0.03, and RMSEA = 0.07), indicating a well-fitting model.

TABLE 3.1: *KMO-MSA and the rotated factor pattern for person-level characteristics for LQ13 learners in SQ13 schools.*

Indicator	KMO-MSA	`Person_F1`: Family structure	`Person_F2`: Learner health	`Person_F3`: Supported retention
`Maternal participation`	0.60	**85***	-1	1
`Co-residence with parents`	0.67	**85***	1	1
`Maternal vitality status`	0.60	**63***	-4	1
`Paternal participation`	0.65	**59***	2	-2
`Good hygiene`	0.62	0	**80***	6
`Good memory`	0.65	-1	**76***	2
`Good communication`	0.68	0	**73***	0
`Retention`	0.56	2	9	**84***
`School feeding`	0.57	4	5	**80***
`Walk to school`	0.70	-4	-3	**60***

The three identified factors represent unique aspects of person-level characteristics:

1. `person_F1` captures variance in family structure, clustering indicators related to the presence and involvement of parents and guardians in the learner's life.

2. `person_F2` clusters indicators describing learners' physical capabilities, focusing on diverse aspects of physical and cognitive function. These encompass abilities such as effective

communication, self-care skills (including dressing and hygiene), and cognitive functions like memory and concentration.

3. `person_F3` groups indicators such as learner access to educational transport, the availability of meals through school feeding programs, and learner retention rates, as a latent factor describing a learner's retention in the system through institutional support.

The overall variance (R^2) explained by the person factors is 56.97%, indicating that just over half of the total variability in the observed variables is accounted for by the underlying person-related latent constructs. This suggests a moderately strong factor solution. The distribution of this explained variance across the three factors is presented in Table 3.2.

TABLE 3.2: *Variance explained by the extracted person factors for LQ13 learners in SQ13 schools.*

Factor	R^2
`Person F1:` Family structure	22.10%
`Person F2:` Learner health	17.60%
`Person F3:` Supported retention	17.20%

In Table 3.3, the Cronbach's alpha values for the extracted person factors are presented. These reliability coefficients indicate the internal consistency of the indicators within each factor, demonstrating that the grouped indicators are sufficiently correlated to measure the same underlying construct. The three person factors exhibit moderate Cronbach's alpha values, suggesting adequate internal consistency.

TABLE 3.3: *Cronbach's alpha values for the extracted person factors for LQ13 learners in SQ13 schools.*

Factor	Cronbach's alpha
`Person F1:` Family structure	0.71
`Person F2:` Learner health	0.65
`Person F3:` Supported retention	0.62

The process is repeated for indicators pertaining to the characteristics of the learner's household, covering aspects such as social security, economic activities, household information, and indicators of welfare and hunger. EFA was performed on a predefined set of indicators, specified in Table A.2 of Appendix A. The suitability of the data for EFA was confirmed by an overall KMO-MSA of 0.69, indicating sufficient common variance among the variables to justify the application of FA. In Table 3.4, the relevant indicators, their corresponding KMO-MSA values, and the factor loadings for the five factors derived from the analysis are presented. CFA results yielded ideal fit indices (CFI = 0.97, SRMR = 0.05, and RMSEA = 0.05), indicating a well-fitting model.

The five identified factors represent unique aspects of household-level characteristics:

1. `hhold_F1` clusters indicators related to food security, a fundamental human right, reflecting a household's ability to consistently access safe, nutritious food to meet dietary needs and preferences.

2. `hhold_F2` groups indicators related to household sanitation, recording access to safe drinking water, sanitation facilities, and waste management.

3. `hhold_F3` describes a household's energy access, clustering indicators for the availability and type of energy sources used.

Table 3.4: *KMO-MSA and the rotated factor pattern for household-level characteristics of LQ13 learners in SQ13 schools.*

Indicator	KMO-MSA	`hhold_F1` Food security	`hhold_F2` Sanitation	`hhold_F3` Energy	`hhold_F4` Welfare income	`hhold_F5` Environment
`Food available`	0.84	**93***	2	2	-3	9
`Food sufficient`	0.84	**92***	2	2	-3	9
`Food has variety`	0.85	**88***	1	2	-3	8
`Household not hungry`	0.85	**87***	1	4	-2	9
`General toilet available`	0.85	-2	**93***	-3	-6	-1
`Managed waste collection`	0.67	0	**90***	-4	-4	3
`Inhouse water`	0.83	6	**82***	9	-8	4
`Main electricity meter`	0.60	6	3	**91***	-1	2
`Energy access`	0.66	1	8	**83***	-1	-2
`Paid electricity`	0.72	1	-9	**78***	2	7
`Grant received`	0.54	1	0	-2	**95***	-2
`Childcare grant received`	0.54	1	3	-4	**94***	-3
`Unearned income earned`	0.90	-9	-18	4	**55***	3
`No air pollution`	0.90	0	-2	9	0	**81***
`No water pollution`	0.70	14	8	0	2	**80***
`No littering`	0.78	13	1	1	-2	**74***

4. `hhold_F4` describes welfare income, derived from state-provided support such as child support grants, which is a significant factor in sustaining households facing economic hardship.

5. `hhold_F5` describes the environment in which the household is located, clustering indicators for pollution and litter.

The overall variance (R^2) explained by the household factors is 73.9%, indicating that a substantial proportion of the total variability in the observed household-related variables is captured by the five underlying latent constructs. This suggests a strong factor solution, reflecting well-defined and distinct dimensions within the household context. The distribution of the explained variance across the five factors is presented in Table 3.5.

Table 3.5: *Variance explained by the extracted household factors for LQ13 learners in SQ13 schools.*

Factor	R^2
`Hhold F1`: Food security	20.51%
`Hhold F2`: Sanitation	14.98%
`Hhold F3`: Energy	13.47%
`Hhold F4`: Welfare income	13.20%
`Hhold F5`: Environment	11.74%

Table 3.6 presents the Cronbach's alpha values for the extracted household factors. The five household factors exhibit high Cronbach's alpha values, suggesting strong internal consistency.

3.5.2 Factor regression for LQ13 learners in SQ13 schools

MRA was employed on the eight extracted factors that were revealed through FA. The composite factors were consolidated as described in §3.4.2 to investigate the relationships among these major themes within the dataset. In Table 3.7, the regression coefficients derived from the

TABLE 3.6: *Cronbach's alpha values for the extracted household factors for LQ13 learners in SQ13 schools.*

Factor	Cronbach's alpha
`Hhold F1:` Food security	0.92
`Hhold F2:` Sanitation	0.87
`Hhold F3:` Energy	0.79
`Hhold F4:` Welfare income	0.76
`Hhold F5:` Environment	0.70

MRA are presented. A regression coefficient was considered statistically significant at a p-value less than 0.05 (*), with particularly strong significance indicated at $p < 0.001$ (**). The colour gradient within each cell visually represents the magnitude of the regression coefficient, with darker shades indicating stronger associations.

TABLE 3.7: *Regression matrix for the eight composite factors based on MRA for LQ13 learners in SQ13 schools. Rows represent predictors, columns represent outcomes. Cell shading indicates the magnitude of the regression coefficient. ** indicates $p < 0.001$, * indicates $p < 0.05$.*

	Family structure	**Supported retention**	**Learner health**	**Food security**	**Welfare income**	**Sanitation**	**Energy**	**Environment**
Family structure	1.00	0.04	-0.00	0.00	0.02	0.09**	-0.08**	0.02
Supported retention	0.04	1.00	0.28**	-0.02	0.28**	-0.08	0.02	-0.02
Learner health	-0.03	0.02	1.00	0.06	0.03	-0.14	-0.03	0.10
Food security	0.00	-0.02	0.01	1.00	-0.03	0.02	0.08*	0.22**
Welfare income	0.02	0.05**	0.00	-0.03	1.00	-0.07**	-0.04	-0.01
Sanitation	0.06**	-0.02	-0.02	0.02	-0.07**	1.00	-0.02	0.05*
Energy	-0.10**	0.02	-0.02	0.03*	-0.04	-0.01	1.00	0.04**
Environment	0.02	-0.01	0.02	0.18**	-0.01	0.05*	0.12**	1.00

Through MRA, a series of significant associations were uncovered, each providing valuable insights into the interconnectedness of the factors within the study context. Expanding on the possible reasons behind these relationships enhances the understanding of the underlying dynamics.

Supported retention positively predicts learner health ($r = 0.28$, $p = 0.001$), suggesting the important role that comprehensive school support programmes play in fostering learner well-being. Schools that provide health services and nutritional support create a more conducive learning environment, which can lead to better academic outcomes. This relationship might also reflect the broader influence of such programmes in reducing absenteeism due to health-related issues, thereby improving retention rates.

Supported retention strongly predicts welfare income ($r = 0.28$, $p < 0.0001$), and the relationship is significantly bidirectional. This suggests that learners from households receiving financial support mechanisms are more likely to attend schools that offer institutional support. Additionally, it could reflect targeted efforts of welfare programmes to align with educational support services, leading to increased retention.

Food security is a strong predictor of the environment ($r = 0.22$, $p < 0.0001$), and this relationship is significantly bidirectional. This suggests that learners from households with adequate and nutritious food are also likely to reside in cleaner areas. Furthermore, the interplay between food security and environmental quality may contribute to a cycle where access to proper nutrition and a healthy living environment mutually reinforce one another, promoting overall well-being.

Environment positively predicts food security ($r = 0.18$, $p < 0.0001$). This points to the interconnectedness of environmental health and nutritional support. Investments in environmental improvements, such as enhancing air and water quality, often go hand-in-hand with improvements in food security.

These regression coefficients suggest that the eight composite factors are closely linked, with each composite factor influencing and being influenced by others. The findings point to the value of adopting holistic and integrated approaches when designing policies and interventions aimed at improving education.

3.5.3 Factor mapping for LQ13 learners in SQ13 schools

SEM was applied to the composite factors comprising LQ13 learners attending SQ13 schools. The results are presented in Table 3.8. These findings provide a useful perspective on the relationships and dynamics within the system, showing how the composite factors interact and influence one another. The outcome of the SEM yielded ideal fit indices (CFI = 0.99, SRMR = 0.0015, and RMSEA = 0.006), indicating a well-fitting model.

TABLE 3.8: *Standardised effects of predictors on outcomes, including the estimated path coefficients (β), standard errors (SE), and t-values for each relationship among the eight composite factors.*

Predictor	Outcome	β	SE	t-value
Environment	Food security	0.21	0.02	11.73
Welfare income	Supported retention	0.13	0.02	6.83
Environment	Energy	0.09	0.02	4.69
Supported retention	Learner health	0.08	0.02	4.26
Family structure	Sanitation	0.08	0.02	4.13
Sanitation	Welfare income	-0.08	0.02	-4.05
Sanitation	Environment	0.05	0.02	5.56
Family structure	Energy	-0.05	0.02	-2.94

The SEM findings and MRA results provide a broad view of the interrelationships among various composite factors. The strongest relationship observed in the SEM results is between the environment and food security ($\beta = 0.21$, $SE = 0.02$, $t = 11.73$), which aligns with the MRA results. This suggests that as families' ability to move to cleaner neighbourhoods increases, their food security is also likely to improve. The environment is also positively associated with energy access ($\beta = 0.09$, $SE = 0.02$, $t = 4.69$), indicating that such a move tends to result in better access to basic services and utilities.

A strong relationship is revealed between welfare income and supported retention ($\beta = 0.13$, $SE = 0.02$, $t = 6.83$), and to a lesser extent, supported retention positively influences learner health ($\beta = 0.08$, $SE = 0.02$, $t = 4.26$). This suggests that programmes supporting learner retention may yield benefits that extend beyond academic outcomes, indirectly supporting learners physically and socially. It also indicates that families receiving financial support are more likely to enrol their children in schools with higher levels of institutional support.

Sanitation shows a negative effect on welfare income ($\beta = -0.08$, $SE = 0.02$, $t = -4.05$), suggesting that improvements in sanitation—as an indicator of better household conditions—may reduce families' reliance on financial support. Conversely, sanitation is positively associated with the household environment ($\beta = 0.05$, $SE = 0.02$, $t = 5.56$), suggesting that improved service delivery may encourage communities to maintain their surroundings. A positive relationship is also observed between family structure and sanitation ($\beta = 0.08$, $SE = 0.02$, $t = 4.13$), indicating that nuclear families are more likely to reside in households with better conditions. A less intuitive finding is the small negative relationship between family structure and energy access ($\beta = -0.05$, $SE = 0.02$, $t = -2.94$).

In Figure 3.2, the relational mapping of the predictors and outcomes included in the estimated path coefficients as a network diagramme of the eight composite factors for LQ13 learners in SQ13 schools is presented.

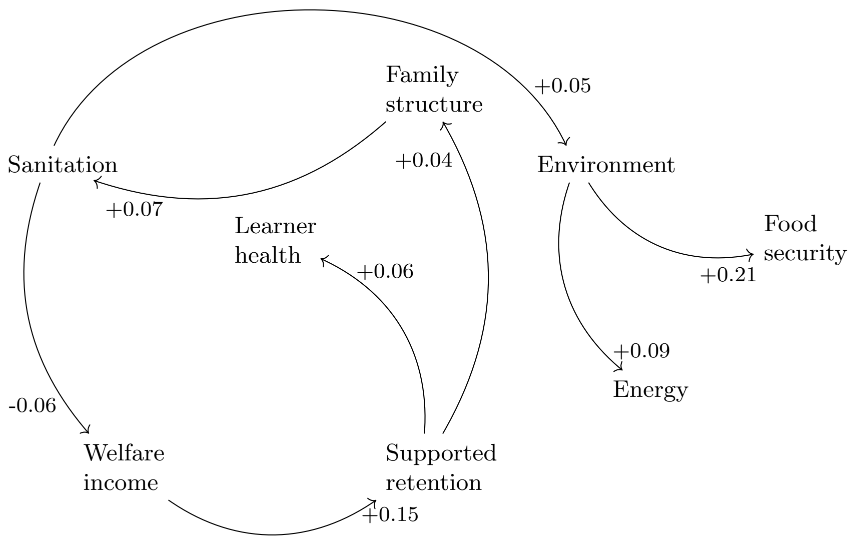

Figure 3.2: *Relational mapping of the predictors and outcomes included in the estimated path coefficients as a network diagrammeme of the eight composite factors for LQ13 learners in SQ13 schools.*

Upon examination of the relational mapping of the predictors and outcomes, it was observed that environment, energy, and food security do not have direct relationships with the learner's education-oriented factors. To further explore how specific indicators within significant composite factors influence learners, an additional analysis was conducted. The five relevant composite factors (*i.e.*, sanitation, family structure, learner health, supported retention, and welfare income) were disaggregated into their original seventeen indicators. MRA was applied to assess the relationships among these indicators, with the resulting regression coefficients presented in Table A.4 in Appendix A. SEM was then employed, and the results are shown in Table A.5 of Appendix A. These relationships, visually represented in Figure 3.3, provide a network-based view of the interconnections among the seventeen indicators for LQ13 learners in SQ13 schools.

In an effort to avoid redundancy, only relationships between indicators that do not originate from the same composite factor are discussed. The strongest of these relationships is observed between the mother's survival status and receipt of a childcare grant ($\beta = 0.12$, $SE = 0.02$, $t = 6.63$). This relationship suggests that the presence of a mother plays an important role in determining eligibility for a childcare grant.

A further association is identified between access to in-house drinking water and the receipt of unearned income ($\beta = -0.13$, $SE = 0.02$, $t = -7.48$). This may suggest that households with access to in-house drinking water are less likely to require unearned income, such as remittances, pensions, or grants.

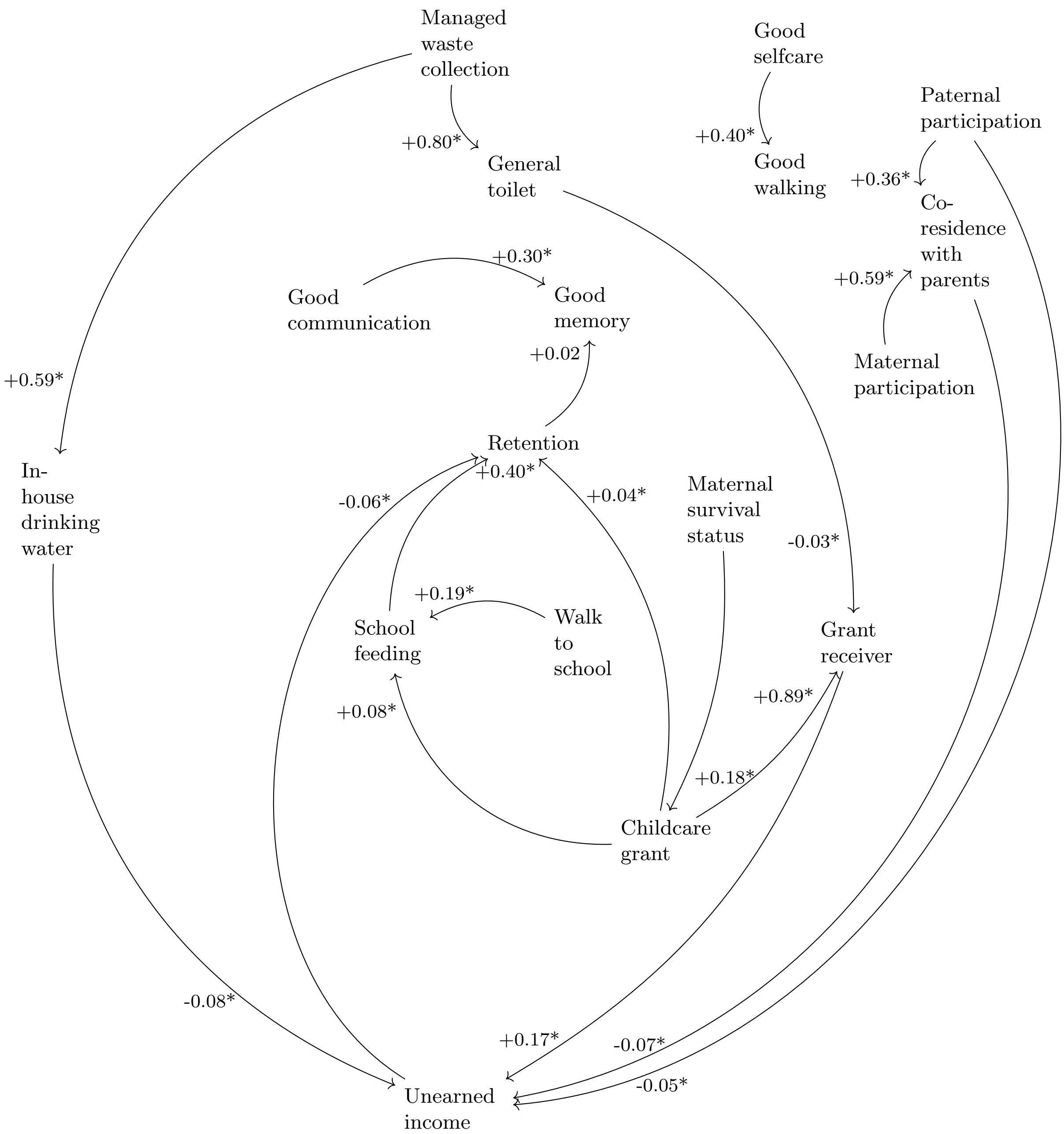

FIGURE 3.3: *Relational mapping of the predictors and outcomes included in the estimated path coefficients as a network diagram of the seventeen indicators that influence LQ13 learners in SQ13 schools.*

Another relationship is observed between co-residence with parents and the receipt of unearned income ($\beta = -0.13$, $SE = 0.02$, $t = -6.55$), suggesting that more secure family structures tend to be more financially stable and, therefore, less reliant on unearned income. The direction of this relationship is worth noting: the absence of unearned income is more likely to follow from co-residence with parents.

Finally, a negative relationship is found between unearned income and retention ($\beta = -0.08$, $SE = 0.02$, $t = -5.15$). This indicates that learners receiving less unearned income are more likely to remain in the school system. Unearned income emerges as a central indicator, linking dimensions such as sanitation, family structure, and supported retention.

To uncover the relationships between factors influencing LQ13 learners in SQ13 schools, a series of analytical techniques was employed. FA was first used to identify eight latent composite factors, providing a broader perspective on the data. These were then analysed using MRA and SEM to examine how the factors relate to and affect one another. This stage of analysis offered insight into the structural patterns in the system. To further investigate the influence of specific indicators on learner retention, a second round of MRA and SEM was conducted on seventeen individual indicators. This level of analysis revealed several associations that are central to the next stage of the study.

3.6 Variate relationship analysis for LQ13 learners in SQ45 schools

The second analysis focused on LQ13 learners attending SQ45 schools, using a sample of 1 716 learners, comprising 53% male and 47% female. Home languages varied, with 25% speaking isiZulu, 18% isiXhosa, 12% Afrikaans, and 11% Setswana. In an effort to determine variate relationships within the South African public high school system for LQ13 learners in SQ45 schools, this dataset is analysed using FA, MRA, and SEM.

3.6.1 Significant factors for LQ13 learners in SQ45 schools

To comprehensively analyse the factors influencing LQ13 learners in SQ45 schools, FA was conducted. Firstly, the analysis centres on variables concerning the person-level characteristics of the learner, including sections on education, health, and family structure. FA was conducted on a predefined set of indicators, which are enumerated in Table A.1 of Appendix A. The suitability of the data for EFA was confirmed with an overall KMO-MSA of 0.74, indicating sufficient commonality among the variables to justify the use of FA. In Table 3.9, the relative indicators, their corresponding KMO-MSA values, and the factor loadings for the three factors discovered are presented. CFA results yielded ideal fit indices (CFI = 0.99, SRMR = 0.03, and RMSEA = 0.03), indicating a well-fitting model.

Table 3.9: *KMO-MSA values and the rotated factor pattern for person-level characteristics of LQ13 learners attending SQ45 schools.*

Indicator	KMO-MSA	Person_F1: Family structure	Person_F2: Learner health	Person_F3: Supported retention
Co-residence with parents	0.57	**89***	1	-2
Maternal participation	0.60	**81***	1	4
Paternal participation	0.65	**72***	2	-11
Good hygiene	0.62	4	**88***	3
Good walking	0.54	3	**87***	0
Good communication	0.79	-1	**52***	-4
School feeding	0.57	-10	2	**80***
Walk to school	0.59	-13	0	**78***
Retention	0.68	11	-4	**57***

The three identified factors, which correspond to unique aspects of person-level characteristics, are comparable to those observed in SQ13 schools. These latent factors represent themes related to family structure, supported retention schemes within schools, and learner health.

The overall variance (R^2) explained by the person factors is 60.18%, indicating that a substantial proportion of the total variance in the observed variables is accounted for by the three extracted factors. This supports the adequacy of the factor structure in capturing key underlying dimensions of person-related influences. The distribution of this explained variance across the three factors is presented in Table 3.10.

TABLE 3.10: *Variance explained by the extracted person factors for LQ13 learners in SQ45 schools.*

Factor	R^2
`Person F1:` Family structure	22.43%
`Person F2:` Learner health	20.15%
`Person F3:` Supported retention	17.60%

In Table 3.11, the Cronbach's alpha values for the extracted person factors are presented. These reliability coefficients assess the internal consistency of the indicators comprising each factor. Family structure demonstrates good internal consistency, learner health shows acceptable reliability, and supported retention yields a lower alpha, indicating moderate internal consistency. Overall, the results suggest that the indicators grouped within each factor are reasonably coherent in measuring the respective underlying constructs.

TABLE 3.11: *Cronbach's alpha values for the extracted person factors for LQ13 learners in SQ45 schools.*

Factor	Cronbach's alpha
`Person F1:` Family structure	0.74
`Person F2:` Learner health	0.65
`Person F3:` Supported retention	0.54

The process is repeated for the indicators pertaining to the characteristics of the learner's household, covering aspects such as social security, economic activities, household information, and indicators of welfare and hunger. EFA was performed on a predefined set of indicators, specified in Table A.2 of Appendix A. The suitability of the data for EFA was confirmed by an overall KMO-MSA of 0.67, indicating sufficient common variance among the variables to justify the application of FA. In Table 3.12, the relevant indicators, their corresponding KMO-MSA values, and the factor loadings for the five factors identified are presented. CFA results yielded ideal fit indices (CFI = 0.97, SRMR = 0.05, and RMSEA = 0.05), indicating a well-fitting model.

The five extracted factors represent distinct dimensions of household-level characteristics that closely align with those identified for SQ13 schools, encompassing food security, welfare income, household sanitation, household energy, and household environmental factors.

The total variance (σ^2) explained by the household factors is 73.88%. This indicates that a considerable portion of the overall variability in the observed household-related indicators is accounted for by the five latent constructs. The result reflects a strong factor solution, with clearly defined and distinct dimensions characterising the household context. The proportion of variance explained by each of the five factors is summarised in Table 3.13.

FA was conducted on learners from two specific strata: Those in the lowest socio-economic quintiles (1–3) attending both low and middle socio-economic schools (Quintiles 1–3 and 4–5). This targeted analysis aimed to capture the learner's perspective within their school context, focusing on factors most relevant to their experience. To improve generalisability, an additional EFA was performed on all learners within each school. Notably, both analyses yielded similar factors describing the respective school systems. This suggests that the factors identified through the specific learner group (Quintiles 1–3) are not solely driven by socio-economic background

Table 3.12: *KMO-MSA and the rotated factor pattern for household-level characteristics for LQ13 learners in SQ45 schools.*

Indicator	KMO-MSA	hhold_F1 Food security	hhold_F2 Sanitation	hhold_F3 Energy	hhold_F4 Welfare income	hhold_F5 Environment
`Food available`	0.76	**93***	7	-7	1	6
`Food sufficient`	0.76	**93***	10	-10	1	6
`Household not hungry`	0.82	**92***	8	-4	1	3
`General toilet available`	0.71	7	**91***	-9	3	0
`Managed waste collection`	0.73	6	**89***	-11	2	1
`Inhouse water`	0.84	9	**82***	-13	13	7
`Main electricity meter`	0.59	2	4	**87***	1	7
`Energy access`	0.67	1	6	**77***	-3	0
`Paid electricity`	0.68	-1	6	**77***	0	4
`Grant receiver`	0.57	-4	-4	0	**95***	-1
`Childcare grant receiver`	0.93	-2	-4	0	**95***	-2
`Unearned income receiver`	0.76	-15	-29	-3	**60***	-5
`No air pollution`	0.72	4	2	5	4	**81***
`No noise pollution`	0.66	-4	0	0	2	**77***
`No littering`	0.58	14	6	-8	4	**74***

but instead reflect broader characteristics of the school environment that influence all learners. In essence, the first analysis provides a focused view of how learners experience their schools, while the second offers a broader account of the school's overall functioning.

Table 3.13: *Variance explained by the extracted household factors for LQ13 learners in SQ45 schools.*

Factor	R^2
`Hhold F1:` Food security	17.49%
`Hhold F2:` Sanitation	16.14%
`Hhold F3:` Energy	14.81%
`Hhold F4:` Welfare income	13.22%
`Hhold F5:` Environment	12.22%

In Table 3.14, the Cronbach's alpha values for the extracted household factors are presented. The reliability coefficients across all five factors suggest a high level of internal consistency, indicating that the grouped indicators are suitably aligned with the constructs they are intended to measure.

Table 3.14: *Cronbach's alpha values for the extracted household factors for LQ13 learners in SQ45 schools.*

Factor	Cronbach's alpha
`Hhold F1:` Food security	0.92
`Hhold F2:` Sanitation	0.87
`Hhold F3:` Energy	0.74
`Hhold F4:` Welfare income	0.81
`Hhold F5:` Environment	0.68

3.6.2 Factor regression for LQ13 learners in SQ45 schools

MRA was employed on the eight composite factors, examining their interrelationships within the dataset. In Table 3.15, the regression coefficients obtained from this analysis are presented. A regression coefficient was considered statistically significant at a p-value less than 0.05 (*), with stronger significance indicated at $p < 0.001$ (**). The colour gradient within each cell visually represents the magnitude of the regression coefficient, with darker shades indicating stronger associations.

TABLE 3.15: *Regression matrix for the eight composite factors based on MRA results for LQ13 learners in SQ45 schools. Rows represent predictors, columns represent outcomes. Cell shading indicates the magnitude of the regression coefficient. ** indicates $p < 0.001$, * indicates $p < 0.05$.*

	Family structure	**Learner health**	**Supported retention**	**Food security**	**Welfare income**	**Sanitation**	**Energy**	**Environment**
Family structure	1.00	0.34	-0.05	-0.01	-0.10**	0.15**	0.00	-0.02
Learner health	0.34	1.00	0.08	-0.11	-0.14	0.06	-0.04	0.32**
Supported retention	-0.05	0.08	1.00	-0.10**	0.38**	-0.19**	0.01	0.01
Food security	-0.01	-0.11	-0.10**	1.00	-0.08**	0.16**	-0.00	0.14**
Welfare income	-0.10**	-0.14	0.38**	-0.08**	1.00	-0.14**	0.00	-0.01
Sanitation	0.15**	0.06	-0.19**	0.16**	-0.14**	1.00	0.28**	0.06
Energy	0.00	-0.04	0.01	-0.00	0.00	0.28**	1.00	0.05**
Environment	-0.02	0.32**	0.01	0.14**	-0.01	0.06	0.05**	1.00

The MRA revealed a series of significant associations between the composite factors, shedding light on the intricate relationships within the study context. By exploring potential reasons for these associations, a deeper understanding of the underlying dynamics is gained.

Welfare income positively predicts supported retention ($r = 0.38$, $p < 0.0001$), and the relationship is significantly bidirectional. Learners who receive support for retention are more likely to be from families receiving welfare income. This might reflect that support programs are effectively targeting learners from lower-income backgrounds.

Energy and sanitation positively predict each other ($r = 0.28$, $p < 0.0001$). Better access to energy is associated with better sanitation. This suggests that areas with more developed energy infrastructure also tend to have better sanitation facilities.

Supported retention negatively predicts sanitation ($r = -0.19$, $p < 0.0001$), and the relationship is significantly bidirectional. Learners receiving institutional support are associated with poorer sanitation conditions. This could indicate that retention support programs are effectively reaching learners in areas with infrastructure challenges.

Food security and sanitation positively predicts each other ($r = 0.16$, $p < 0.0001$). Higher food security is linked to better sanitation conditions. This might indicate that areas with better food access also have better overall infrastructure and living conditions.

Family structure and sanitation positively predict each other ($r = 0.15$, $p < 0.0001$). Learners from more secure family units are more likely to have access to better sanitation. This might reflect that secure family structures may live in areas with better infrastructure or have more resources for maintaining sanitary conditions.

3.6.3 Factor mapping for LQ13 learners in SQ45 schools

SEM was employed on the dat and the results of this analysis are shown in Table 3.16. These findings provide valuable insights into the complex relationships and dynamics within the system.

TABLE 3.16: *Standardised effects of predictors on outcomes including the estimated path coefficients (β), standard errors (SE), and t-values for each relationship of the composite factors in the structural equation model for LQ13 learners in SQ45 schools.*

Predictor	Outcome	β	SE	t-value
Welfare income	Supported retention	0.33	0.02	15.54
Sanitation	Welfare income	-0.21	0.02	-9.45
Family structure	Sanitation	0.19	0.02	8.34
Sanitation	Supported retention	-0.16	0.02	-7.00
Sanitation	Food security	0.15	0.02	6.38
Family structure	Welfare income	-0.13	0.02	-5.45
Sanitation	Energy	0.13	0.02	5.30
Environment	Food security	0.12	0.02	4.97
Welfare income	Food security	-0.11	0.02	-4.51
Food security	Supported retention	-0.09	0.02	-4.09
Energy	Environment	0.10	0.02	4.02

The SEM analysis reveals a robust positive regression coefficient between welfare income and supported retention ($\beta = 0.33$, $SE = 0.02$, $t = 15.54$). This finding suggests that families in receipt of welfare income are more likely attend a school that provides institutionalised support. The practical implications underscore the efficacy of social welfare programmes in promoting educational retention, indicating that financial support plays a pivotal role in enabling continued education for children from lower-income families.

A noteworthy positive relationship is observed between family structure structure and sanitation ($\beta = 0.19$, $SE = 0.02$, $t = 8.34$). This implies that more secure family structures may be more inclined to invest in improved sanitation facilities, possibly due to more focused resource allocation within the household.

Interestingly the SEM analysis unveils several negative relationships. Fristly, a negative relationship between sanitation and supported retention ($\beta = -0.16$, $SE = 0.02$, $t = -7.00$), suggesting that areas with better sanitation may experience lower institutionalised support schemes from the schools. Further, a negative relationship between sanitation and welfare income ($\beta = -0.21$, $SE = 0.02$, $t = -9.45$), indicating that areas with better sanitation may have lower levels of welfare income. This could be because improved living conditions are associated with reduced reliance on welfare support. A negative relationship between family structure structure and welfare income ($\beta = -0.13$, $SE = 0.02$, $t = -5.45$), suggesting that nuclear families may be less likely to receive welfare income, possibly due to increased household stability or economic self-sufficiency. A negative relationship is revealed between food security and supported retention ($\beta = -0.09$, $SE = 0.02$, $t = -4.09$), this describing that learners who are food insecure are not receiving institutionalised support. Although this relationship may initially appear counterintuitive, it indicates that LQ13 learners in SQ45 schools are more likely to experience food insecurity due to a reduced provision of support programmes, such as school meals. Finally, a negative relationship is present between welfare income and food security ($\beta = -0.11$, $SE = 0.02$, $t = -4.51$), revealing that households in need of welfare income are generally less likely to be food secure.

An intriguing relationship is observed between sanitation and food security ($\beta = 0.15$, $SE = 0.02$, $t = 6.38$), as well as sanitation and energy access ($\beta = 0.13$, $SE = 0.02$, $t = 5.30$), showing the interconnectedness of a clean environment, energy access, and access to nutritious food.

A significant positive relationship is observed between the environment and food security ($\beta = 0.12$, $SE = 0.02$, $t = 4.97$). This suggests that a better environment contributes to improved food security. Finally, the SEM analysis identifies a positive relationship between energy access and environmental outcomes ($\beta = 0.10$, $SE = 0.04$, $t = 4.02$). This suggests that access to energy may lead to better environmental outcomes.

Overall, the SEM results largely validate the findings of the MRA, with some variations in the strength of the relationships. The SEM provides a more interconnected view of the factors at play, allowing for a more nuanced understanding of how these variables influence one another within a complex system. The consistency between the two methods increases confidence in the identified relationships. In Figure 3.4, the relational mapping of the predictors and outcomes included in the estimated path coefficients as a network diagrammeme of the eight composite factors for LQ13 learners in SQ45 schools is presented.

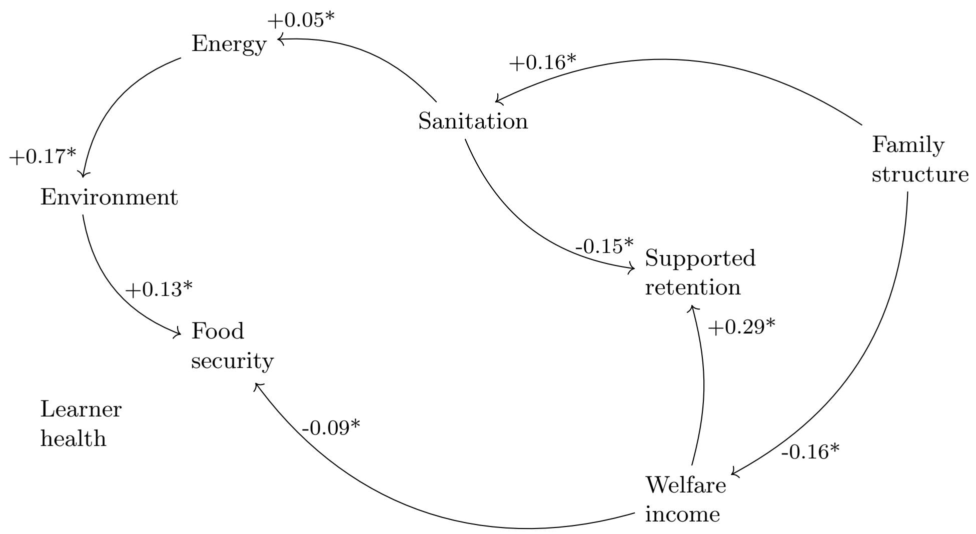

FIGURE 3.4: *Relational mapping of the predictors and outcomes included in the estimated path coefficients as a network diagram of the eight composite factors for LQ13 learners in SQ45 schools.*

An in-depth examination of the relational mapping of the predictors and outcomes, revealed that the composite factors of environment, energy access, and food security do not directly have relationships with educational factors of the learner. Additionally, learner health appeared to be isolated, with no significant connections to other factors. To gain a more granular understanding of the relationships between individual indicators and learner outcomes, a supplementary analysis was conducted. This analysis focused on the indicators that were part of the composite factors found to have a significant influence on learners, namely sanitation, family structure, supported retention, and welfare income.

MRA was employed on these twelve indicators to evaluate their relationships and understand how they interact and connect with each other. This analysis aimed to provide insights into the specific individual indicators that contribute to the overall influence of the composite factors on learners. In Table A.6 of Apendix A, the regression coefficients derived from the MRA are presented.

To further explore the relationships between individual indicators, SEM was applied to the dataset comprising the twelve indicators. The results of this analysis are presented in Table A.7 in Appendix A. For a visual representation of the relationships identified through SEM, Figure 3.3 presents a network diagramme depicting the estimated path coefficients among the twelve indicators for LQ13 learners in SQ45 schools. This diagramme provides a clear visualisation of the complex interplay between these variables. The outcome of the SEM yielded ideal fit indices

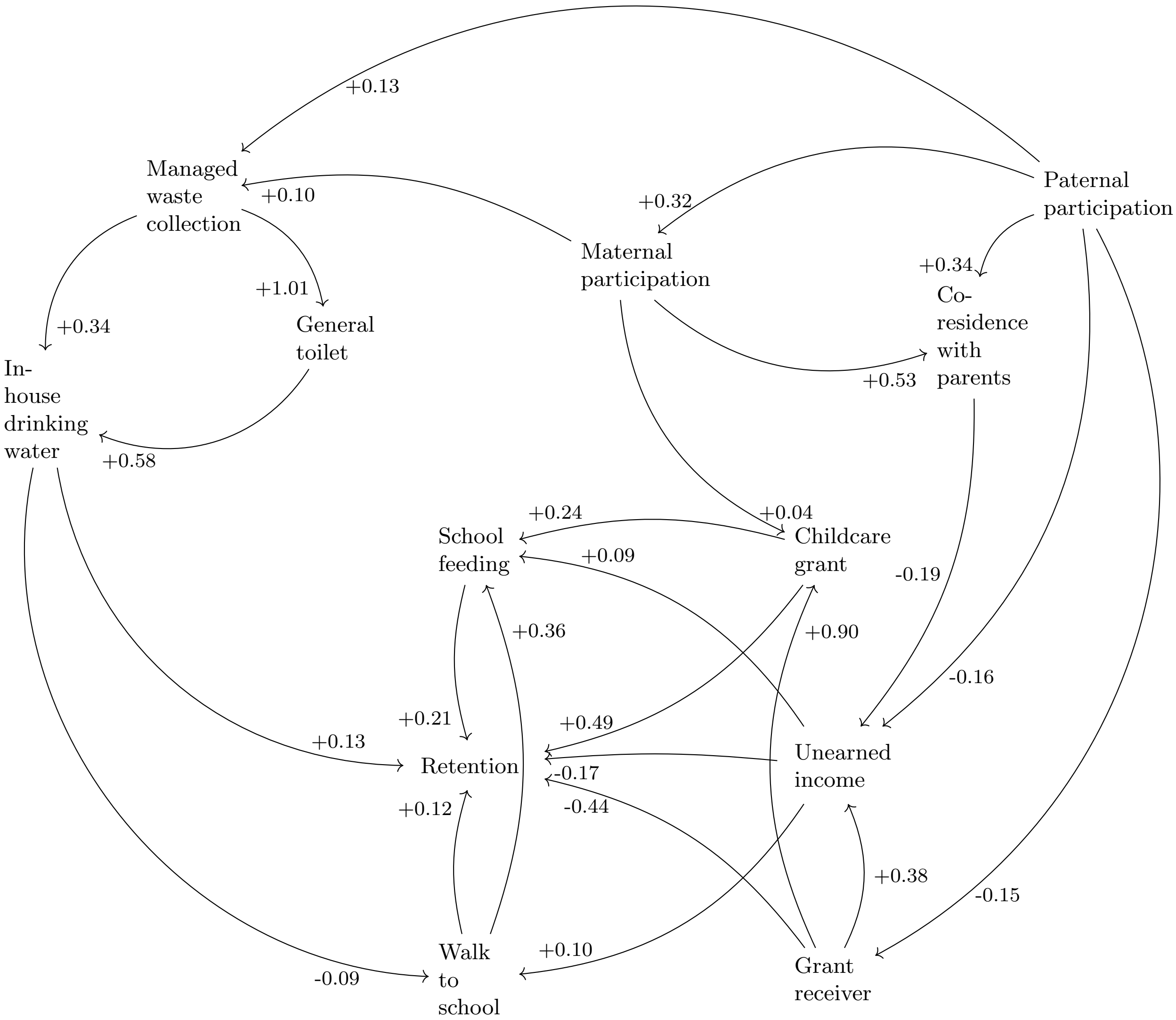

Figure 3.5: *Relational mapping of the predictors and outcomes included in the estimated path coefficients as a network diagram of the twelve indicators that influence LQ13 learners in SQ45 schools.*

In an effort to avoid redundancy and repetition, again only relationships between indicators that do not originate from the same composite factor are discussed. The strongest of these relationships is observed between the childcare grant and retention ($\beta = 0.49$, $SE = 0.05$, $t = 10.05$). This suggests that if a learner is retained in the school, they are highly likely to also be a recipient of a childcare grant.

A counterintuitive finding is the negative relationship between grant receipt and retention ($\beta = -0.44$, $SE = 0.05$, $t = -8.91$). This is particularly interesting because, despite the high regression coefficient between childcare grant receipt and overall grant receipt, which could imply that they represent similar constructs—they exhibit opposing effects on retention rates. A plausible interpretation is that receiving a grant, in general, is insufficient to ensure school retention; it specifically needs to be a childcare grant.

Another notable positive relationship is identified between the receipt of childcare grants and the participation in school feeding schemes ($\beta = 0.24$, $SE = 0.02$, $t = 11.25$). This may indicate that children attending schools that offer feeding schemes are also likely to receive a childcare grant. This finding shows the notion that learners from households reliant on welfare income tend to enrol in schools that provide institutionalised support programmes.

Lastly, the negative relationship between co-residence with parents and the receipt of unearned income ($\beta = -0.19$, $SE = 0.02$, $t = -7.83$) is worth noting. This relationship suggests that learners who live in stable family structures are less likely to receive unearned income such as remissions, pensions, or grants.

In an effort to derive meaningful relationships that represent LQ13 learners in SQ45 schools, a series of analytical techniques was employed. FA identified eight latent composite factors, which offered a broader perspective on the data. MRA and SEM were applied to these eight composite factors to better understand their interrelationships and the influence they exert on one another. These analyses provided valuable insights into the complexities involved in understanding schooling systems. Furthermore, to attain a more detailed understanding of the specific indicators influencing learner retention, a second round of MRA and SEM was conducted on twelve indicators. This revealed intriguing relationships that are crucial to the continued analysis in this study.

3.7 Learner profile comparison

The characteristics defining lower socio-economic learners in both the SQ13 and SQ45 schools have been identified. To deepen the understanding of these learners, a thematic value was assigned to each learner. By considering each binary response in the encoded data and multiplying it by the respective relative factor loadings, then dividing by the total factor loadings, a composite initial factor score was derived for each learner as depicted in Equation 3.4.

$$F_i = \frac{\sum_j^n \ell_j \iota_j}{\sum_j^n \ell_j}, \tag{3.4}$$

where ℓ_j is the factor loading for indicator j derived from FA, and ι_j is the binary encoding of the indicator j, for the n indicators that comprise each composite factor. Averaging these scores across all LQ13 learners in their respective schools allowed for an overall assessment of each group as expressed in Equation 3.5.

$$\bar{F_k} = \frac{1}{m} \sum_{i=1}^{m} F_i \quad \forall k, \tag{3.5}$$

where m is the number of observations.

This approach facilitates the identification of the defining similarities between these two groups of learners, as well as the specific areas in which they diverge. The average scores derived from this method can be expressed as "goodness scores," where values closer to one indicate a healthier or

Table 3.17: *Resultant factor themes "goodness scores" for LQ13 learners in SQ13 and SQ45 schools respectively*

Factor ($\bar{F}_k$)	SQ13	SQ45
Family structure	0.58	0.57
Learner health	0.99	0.99
Supported retention	0.92	0.53
Food security	0.70	0.84
Welfare income	0.61	0.38
Sanitation	0.36	0.65
Energy	0.94	0.95
Environment	0.76	0.82

more advantaged learner, and scores closer to zero reflect learners who are comparatively worse off. These scores are summarised in Table 3.17.

Analysis of these initial average factor scores reveals notable differences between the two systems in terms of supported retention, welfare income, and sanitation. LQ13 learners who attend LQ13 schools benefit from increased levels of supported retention and more frequent welfare income from the government. This trend is anticipated, as LQ13 schools typically implement feeding programmes, thereby contributing to a higher supported retention factor score.

Interestingly, welfare income demonstrates considerable variation despite the learners belonging to the same socio-economic class. This suggests that LQ13 learners in SQ13 schools may be encouraged to apply for grants, whereas those in SQ45 schools might be on scholarships, which may reduce their need for welfare income. In contrast, LQ13 learners attending SQ45 schools appear to have access to improved sanitation at home. This disparity could be attributed to these learners often residing near more affluent schools, which are typically situated in wealthier areas, thus affording them access to better sanitation facilities.

The "goodness scores" offer a meaningful perspective on the learners' overall status, providing a comparative measure between SQ13 and SQ45 schools. By integrating the factor themes derived from FA with the factor relationships identified through MRA and SEM, these scores highlight both the distinct disparities and the commonalities between the two school groups. This approach yields a comprehensive understanding of the learner profiles within each school type, illustrating key areas of advantage and disadvantage across socio-economic contexts.

3.8 Chapter 3 summary

In this chapter, the relationships between variables within the context of the South African education system were explored. The chapter commenced with an introduction to the GHS. Following a comprehensive examination of the dataset, the initial phase focused on identifying patterns, trends, and potential associations within the data. Thereafter, a detailed account of the data selection and preparation procedures employed in the study was presented. The theoretical framework underpinning the methodologies used to investigate these relationships was then outlined. The analytical techniques employed included FA, MRA, and SEM. The application of these methods to two distinct datasets was subsequently discussed. Lastly, the learner profile was unpacked.

CHAPTER 4

Factor importance in the South African public high school system towards a survey design

Contents

In this chapter, factor importance among the variables identified as characterising lower socio-economic learners within two distinct schooling systems in South Africa was determined, and a survey was designed in pursuit of Objective III. Section 4.1 outlined the significance of factor importance within the context of the study, serving as a means of applying the results obtained in the previous chapter. Thereafter, Section 4.2 introduced the theoretical framework of ML, followed by an explanation of the model pipeline in Section 4.3. This pipeline was subsequently applied to two separate datasets, namely LQ13 learners within SQ13 and SQ45 schools, in Section 4.4 and Section 4.5, respectively. Finally, Section 4.6 considers a survey design.

4.1 Factor importance as a foundation for survey design

In Chapter 3, the study identified the factors and interrelationships that characterise lower socio-economic learners across two distinct educational systems. Drawing on the research discussed in §2.3, by Slamang [83] and Becker [12], valuable insights have emerged. Slamang argued that enhancing learner progression requires focusing on the classroom environment, particularly in relation to class size, school resources, and teacher effectiveness. Becker, on the other hand, emphasised that although the classroom environment influences learners, the importance of personal circumstances, especially the household conditions in which they reside, should not be overlooked. She contends that the family environment also plays a crucial role in shaping learner progression.

The contrast between these perspectives does not reduce the relevance of either approach; rather, it reflects the nature of the data used in each study. Slamang's dataset is skewed towards the school environment, comprising mostly school-relevant data. Conversely, the 2019 GHS dataset used in Becker's study includes limited classroom-related variables and focuses mainly on household characteristics.

In this study, the findings suggest that sanitation, family structure, supported retention, and welfare income are significant influences on learner retention. While supported retention emerges as a notable factor, household-related aspects are more prominent. This prominence can be attributed to the structure of the 2022 GHS dataset, which places considerable focus on household conditions and offers limited insight into the quality of education. Although some school-related variables appeared in Becker's study, most of the classroom environment questions found in the 2019 GHS are missing from the 2022 edition. For findings to be practically applied to a case study school, the comprehensiveness of the data is important. The dataset should provide a balanced representation of the learner's personal traits, household context, and school environment.

To move from the factors and interrelationships identified for lower socio-economic learners in two schooling systems to a practical case study, new and balanced data must be collected. Ensuring that this data avoids bias towards household or school-related variables requires a carefully designed survey. This includes thoughtful structuring and wording of the questions to achieve a representative view of all relevant dimensions.

Respondent fatigue refers to a decline in response quality as a survey continues, often due to tiredness, boredom, or cognitive overload [48]. Given the evidence supporting this concern, and acknowledging that a lengthy questionnaire may overwhelm learners, the number of questions posed is deliberately limited. To assist in this refinement, ML techniques are applied to evaluate the relative importance of each factor theme in predicting learner progression. This process ensures that only the most relevant factors are retained in the final questionnaire.

4.2 Theoretical framework of machine learning

Learning, as a general process, involves acquiring new or altering existing behaviours, values, knowledge, skills, or preferences. Unlike humans, who learn naturally through experience, machines rely on data to facilitate learning [3].

At its core, ML is a widely applied subfield of artificial intelligence that enables computers to learn and improve performance without being explicitly programmed [65]. The aim is to optimise decision-making by allowing algorithms to adjust their outputs based on patterns identified in data, with accuracy measured by how often the predictions or classifications are correct.

ML techniques can uncover novel associations, generate new hypotheses, and drive research towards more informed and effective decisions. The term "machine learning" was first coined by Arthur Samuel in 1959 to describe the ability of computers to learn from data without being explicitly programmed [76].

Within ML, six broad paradigms describe how learning can occur: Supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning, transduction, and learning to learn [64]. This study focuses on supervised learning, which uses labelled data to train models that map inputs to known outputs. In this context, the goal is to predict an educational outcome, specifically, learner progression through the school system [49].

Supervised learning algorithms are typically grouped into the two main categories of classification and regression. Classification algorithms predict discrete labels or categories. For example, a model may classify a learner as either *high risk* or *low risk* of dropping out based on factors like attendance, academic performance, and socio-economic status [57]. Regression algorithms, on the other hand, predict continuous outcomes. For instance, regression might be used to estimate the degree of improvement in a learner's academic performance over time based on a set of predictor variables [42].

Table 4.1 provides an overview of commonly used ML techniques, summarising each technique's core functionality and examples of relevant applications in education and related fields.

TABLE 4.1: *Overview of commonly used ML techniques.*

Technique	**Description**	**Literature**
Linear regression	Employed for regression tasks, this method aims to predict continuous values by establishing a linear relationship between independent and dependent variables.	[2], [38]
Logistic regression	Used for binary classification problems, this approach estimates the probability of an event occurring through the application of a logistic function.	[21], [68]
Support vector machines	Used for classification and regression purposes, support vector machines identify the optimal hyperplane that effectively separates classes within the feature space.	[79], [44]
Naive Bayes	This is a probabilistic algorithm employed for classification that assumes independence among features within a given class.	[80], [59]
k-nearest neighbors	Used for both classification and regression, the k-nearest neighbours algorithm classifies new instances based on the majority vote of its k-nearest neighbours.	[5], [45]
Random forest	An ensemble learning method, this approach combines multiple decision trees to enhance the accuracy and robustness of predictions.	[8], [75]
Gradient boosting	An ensemble learning method, this technique integrates multiple weak models to formulate a robust predictor, thereby rectifying the errors of preceding models.	[13], [74]

4.3 Factor importance analysis model pipeline

The model pipeline used to conduct factor importance analysis within the South African public high school system is depicted in Figure 4.1. This pipeline consists of four stages: data preprocessing, selection of the ML technique, grid search for optimal parameters, and an iterative approach to establish the best model configuration.

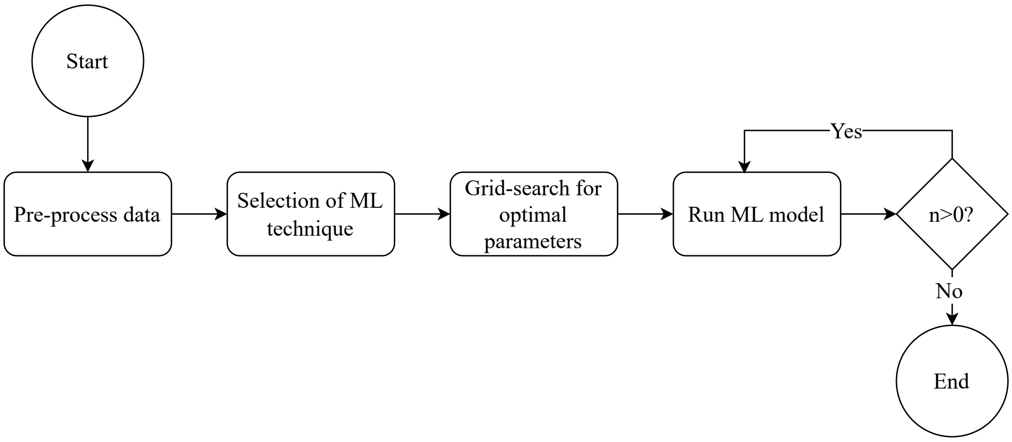

FIGURE 4.1: *Pipeline for factor importance analysis including pre-processing, model tuning, and iterative factor reduction for n factors, simplified from [40].*

This study's ML component aimed to predict whether a learner would pass or fail a grade, identifying the key factors affecting grade progression. Using 2022 GHS data, the model incorporated factors such as province, grade, age-grade delta, gender, home language, and the indicator variables derived from FA. The target variable was the learner's pass status in 2021, retrospectively inferred based on whether the learner had to repeat their grade. This allowed learners to be categorised as either having passed or failed.

A notable challenge was the class imbalance in the target variable, as approximately 85% of learners reported having passed. This posed a risk of biasing the model towards the majority class. To mitigate this, the Synthetic Minority Over-sampling Technique (SMOTE), developed by Chawla *et al.*, was employed [26]. SMOTE generates synthetic samples for the minority class (learners who failed or dropped out), improving the model's capacity to predict rare outcomes and enhancing its overall classification performance.

Van der Heever *et al.* [97] conducted a study on South Africa's high school system using a combination of ML and agent-based modelling, based on the 2019 GHS. Among the tested ML methods, gradient boosting showed the highest predictive accuracy, achieving an error rate of 2.95%. Based on this result, gradient boosting was selected as the technique for the current study.

A grid-search approach was used to determine the optimal hyperparameters for the gradient boosting model. Once fitted, the model was used to assess factor importance, quantifying how much each variable contributed to predictive performance. Factor importance in this context does not indicate a direct relationship with passing or failing. Rather, it reflects how informative a factor is for distinguishing between learners at risk of failing and those likely to succeed.

To determine the minimum number of factors required to preserve model accuracy, an iterative reduction approach was used. In each round, the least important factor was removed, and the model was retrained. This continued until only one factor remained. The process enabled a visualisation of how accuracy changes as factors are eliminated, offering practical guidance for designing a learner questionnaire with only the most essential questions.

4.4 Factor importance for LQ13 learners in SQ13 schools

For the factor importance analysis of LQ13 learners in SQ13 schools, the dataset comprised 2 839 entries and 68 factors. The grid search–selected parameters are listed in Table 4.2. Model performance was evaluated using precision, recall, and the F1-score. As shown in Table 4.3, the gradient boosting model achieved a precision of 0.88 for predicting learners who passed. Precision, which is the proportion of true positives among all positive predictions, reflects the model's accuracy in identifying successful learners. Recall, which measures the model's ability to capture all actual positives, was 0.95. The F1-score, which balances precision and recall, reached 0.92 for 'pass' predictions, confirming the model's strong predictive capability for learner progression.

TABLE 4.2: *Key gradient boosting classifier parameters for LQ13 learners in SQ13 schools*

Parameter	Value
`learning_rate`	0.3
`n_estimators`	200
`max_depth`	4
`min_samples_split`	5
`min_samples_leaf`	2
`loss`	log_loss
`subsample`	1.0
`validation_fraction`	0.1

TABLE 4.3: *Classification metrics by class for LQ13 learners in SQ13 schools*

Class	Precision	Recall	F1-Score
Fail	0.49	0.26	0.34
Pass	0.88	0.95	0.92

A confusion matrix, as depicted in Figure 4.2, provides a performance evaluation for classification models. It outlines the model's predictions against the actual outcomes, categorising results into true positives, true negatives, false positives, and false negatives. This matrix is crucial for assessing a model's strengths and weaknesses, especially in distinguishing between different classes.

The final results revealed an overall prediction accuracy of 85.00%. As depicted in Figure 4.2, the model correctly identified 26.19% of the 'Fail' class and 95.25% of the 'Pass' class. However, it misclassified 73.81% of the negative cases as positive and 4.75% of the positive cases as negative. These findings emphasise the model's robust recall and precision for the 'Pass' class, while also indicating a need for improvement in its handling of the minority 'Fail' class.

Figure 4.3 visually depicts the factor importance obtained through the ML model. To reiterate, factor importance reflects how well a factor contextualises learners, enabling the model to distinguish between those likely to pass or fail. The factor `Learner retention` is the most significant contributor to the model's predictions, indicating that regular school attendance is essential for grade progression. Next, the factor `Age`, which relates to the grade-age delta, underscores the importance of learners being at the appropriate age for their grade level. Lastly, the factor `Child support grant` highlights the importance of financial assistance for learners in lower-quintile schools, suggesting that such support plays a crucial role in their academic success.

The results derived from the iterative approach of sequentially removing the least significant factors are presented in Figure 4.4. The x-axis represents the number of factors, while the y-axis indicates the overall accuracy of the ML model. For a school classified as SQ13, the accuracy remains relatively stable at approximately 80% until the removal of the 16^{th} factor.

By concentrating on the top sixteen factors, it is possible to reduce respondent fatigue while maintaining adequate predictive accuracy. This suggests that by considering only sixteen factors of a learner in an LQ13 context within an SQ13 school, one can predict their likelihood of grade progression with approximately 80% confidence.

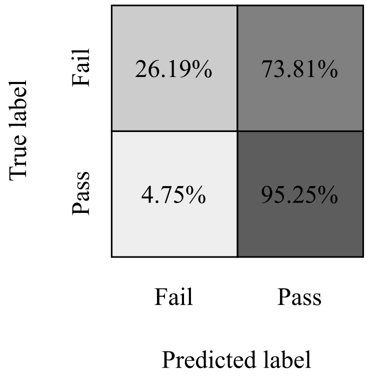

FIGURE 4.2: *Confusion matrix for LQ13 learners in SQ13 schools.*

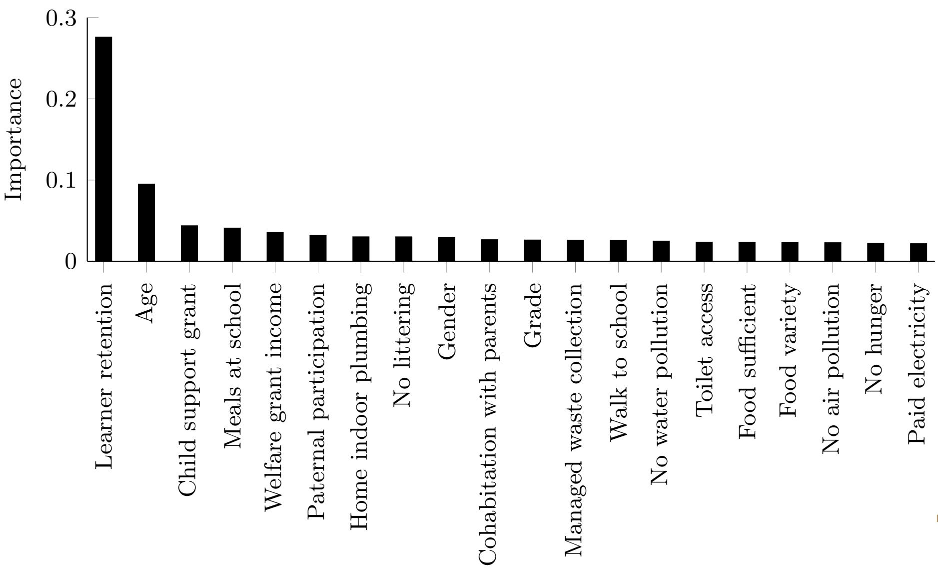

FIGURE 4.3: *Factor importance for LQ13 learners in SQ13 schools.*

4.5 Factor importance for LQ13 learners in SQ45 schools

For the factor importance analysis conducted on LQ13 learners within SQ45 schools, a dataset comprising 1 716 entries and 62 factors was used. Table 4.4 presents the selected parameters from the grid search. The gradient boosting model's performance was evaluated based on precision, recall, and the F1-score. As shown in Table 4.5, the model achieved a precision of 0.93 in predicting learners who passed. The model's recall was recorded at 0.94 for passing learners. The F1-score reached 0.94 for 'pass' predictions.

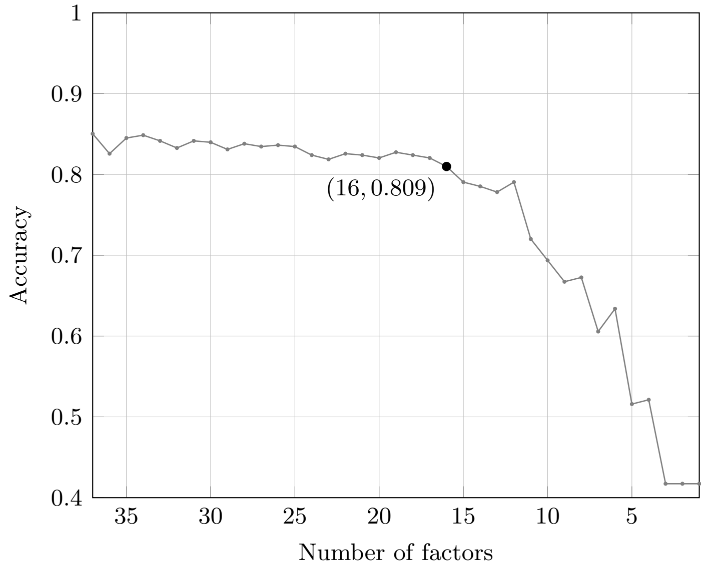

FIGURE 4.4: *Change in learner progression prediction accuracy against number of factors for LQ13 learners in SQ13 schools.*

TABLE 4.4: *Key gradient boosting classifier parameters for LQ13 learners in SQ45 schools*

Parameter	**Value**
`learning_rate`	0.3
`n_estimators`	300
`max_depth`	5
`min_samples_split`	2
`min_samples_leaf`	1
`loss`	log_loss
`subsample`	1.0
`validation_fraction`	0.1

TABLE 4.5: *Classification metrics by class for LQ13 learners in SQ45 schools*

Class	**Precision**	**Recall**	**F1-Score**
Fail	0.51	0.41	0.49
Pass	0.93	0.94	0.94

The final results indicated an overall prediction accuracy of 89.00%. As shown in Figure 4.5, the model accurately classified 46.15% of the 'Fail' class and 94.43% of the 'Pass' class. However, it misclassified 53.85% of the negative instances as positive and 5.57% of the positive instances as negative. These results highlight the model's strong recall and precision for the 'Pass' class, while also suggesting the need for enhanced handling of the minority 'Fail' class.

Figure 4.6 provides a visual representation of factor importance as determined by the ML model. The factor `Learner retention` emerges as the most influential factor in the model's predictions, suggesting that consistent school attendance is crucial for grade progression. The next most significant factor, `Meals at school`, underscores the importance of food provision for learners in higher-quintile schools, indicating that such assistance plays a vital role in supporting academic success. Finally, the factor `Age`, which reflects the grade-age alignment, highlights the importance of learners being of an appropriate age for their respective grades.

The results obtained from the iterative process of sequentially eliminating the least significant factors are shown in Figure 4.7. The x-axis denotes the number of factors, while the y-axis

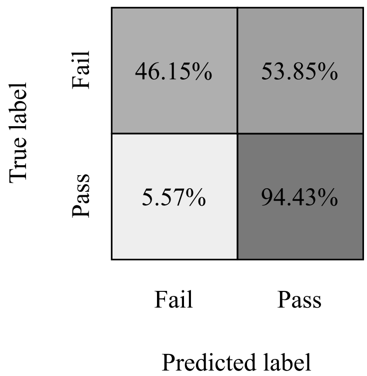

FIGURE 4.5: *Confusion matrix for LQ13 learners in SQ45 schools.*

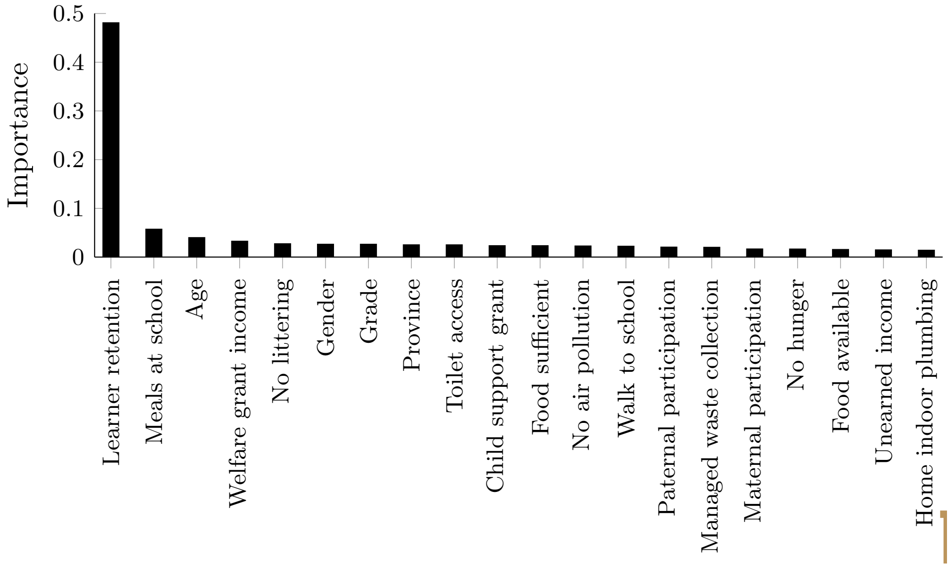

FIGURE 4.6: *Factor importance for LQ13 learners in SQ45 schools.*

represents the overall accuracy of the ML model. For an SQ45 school, accuracy remains consistently high, at approximately 90%, before eventually declining. It is noteworthy that even after the removal of the seventh factor, the model maintains an accuracy exceeding 80%.

4.6 Survey design

Considering the factor importance analysis of LQ13 learners in both SQ13 and SQ45 schools, several potential consequences, effects, and outcomes have emerged as critical within these educational environments. While some similarities exist between the two school systems, there are marked contextual differences in the challenges learners face in their respective environments.

For LQ13 learners in SQ13 schools, consistent attendance and appropriate age-grade placement emerge as the most significant factors influencing academic success. This finding highlights the prevalent issue of grade repetition and over-age learners in South African schools, often

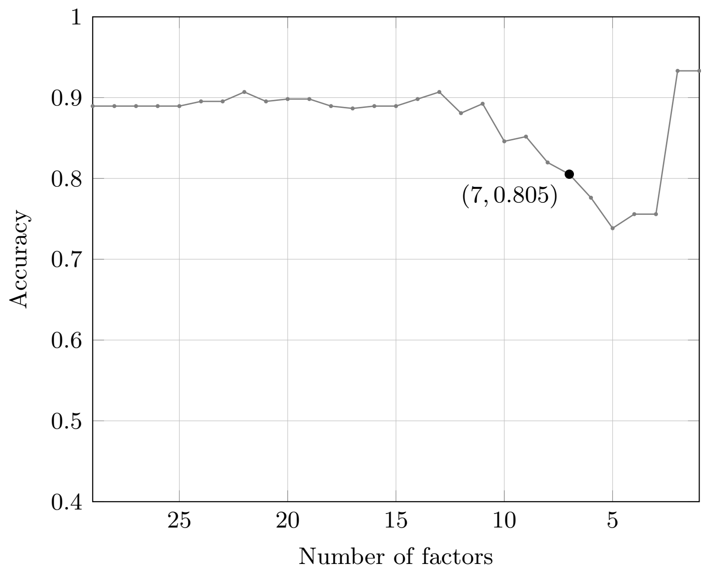

FIGURE 4.7: *Change in learner progression prediction accuracy against number of factors for LQ13 learners in SQ45 schools.*

linked to late school enrolment, repeated grades, or interruptions in education due to socio-economic hardships. The relatively lower ranking of school meals in this context may suggest that community-based mechanisms partly alleviate food insecurity.

In contrast, for LQ13 learners in SQ45 schools, while attendance and age-grade alignment remain important, school meals rank higher in importance. This suggests that learners in these schools may experience greater levels of food insecurity compared to their more affluent peers, making school feeding programmes a critical support.

These findings reinforce the essential role of social support systems in education. There remains a clear need for targeted interventions that promote regular attendance and prevent grade-age misalignment through early support. School feeding programmes prove beneficial across socio-economic contexts and are central to learner well-being and academic success. Additionally, financial grants remain a crucial support mechanism, particularly for learners facing economic constraints.

Based on the iterative approach employed, sixteen core questions were developed. These questions can be administered to any lower socio-economic learner in South Africa and used to predict, with at least 80% accuracy, whether the learner will pass their current grade without additional interventions. The questions are:

1. Do you attend a school that charges fees?
2. What is your gender?
3. How old are you?
4. What grade are you currently enrolled in?
5. Do you walk to school?
6. Does your school provide meals to learners?
7. Is your father actively involved in your life?

8. Do you live with both parents?
9. Does your household have access to clean drinking water?
10. Is there water pollution in your area or household?
11. Is there littering in or around your home?
12. Does your family receive a grant from the government?
13. Does your family receive a child care grant?
14. Is there enough food available in your household to meet your needs?
15. Is your household's waste collected?
16. Do you have a flushing toilet at home?

However, it is recognised that these questions do not capture the quality of education. In the 2019 GHS, questions relating to educational quality were included and found significant in Becker's analysis [12]. These questions were removed in the 2022 GHS. As a result, Becker's earlier findings are used to inform the construction of additional questions to capture educational quality. These include:

1. Do you feel you receive a good quality of teaching at school?
2. Is your teacher frequently absent from school?
3. Does your school have enough teachers to meet learners' needs?
4. Is your class size small?
5. Do you have access to enough learning materials and books at school?
6. Is the school building in good condition?
7. Has your teacher participated in a strike recently?
8. Are you able to read well?
9. Are you able to write your own name?
10. Can you read and understand road signs?
11. Can you fill out a form independently?
12. Can you calculate the correct change when you buy something?
13. Are you able to write a letter to someone?

This survey design serves as a data-informed collection tool, enabling the gathering of statistically significant information. The selected questions are structured to extract meaningful responses from learners in lower socio-economic contexts within South Africa. However, they are broadly applicable to other domains and research settings. Rather than being tailored to a specific school, the questions aim to represent the broader population of learners from socio-economically disadvantaged backgrounds in the South African education landscape. As such, the survey instrument may be adapted to investigate additional research questions that require a holistic yet statistically robust understanding of this learner group, while maintaining key variables and mitigating respondent fatigue.

4.7 Chapter 4 summary

In this chapter, factor importance among the variables that characterised lower socio-economic learners within two distinct schooling systems in South Africa was determined, and a survey was designed in pursuit of Objective III. The significance of factor importance within the context of the study was outlined as a means of applying the findings from the previous chapter as a basis towards a survey design. The theoretical framework of ML was introduced, followed by an explanation of the model pipeline. This pipeline was applied to two separate datasets, namely LQ13 learners within SQ13 and SQ45 schools. Finally, survey design was considered.

CHAPTER 5

Variate analysis of learner progression determinants within the case study school

Contents

This chapter builds on the findings presented in Chapters 3 and 4 with the aim of analysing the determinants of learner progression within an independent high school in South Africa, hereafter referred to as the focus school, in alignment with Objective IV. The analysis begins in Section 5.1, which provides contextual information about the selected school. Section 5.2 summarises the dataset obtained from the school. Variate relationships are subsequently examined using FA, MRA, and SEM, as outlined in Sections 5.3.1, 5.3.2, and 5.3.3, respectively. Finally, ML techniques are applied in Section 5.4 to identify the learner attributes most strongly associated with successful academic progression.

5.1 Introduction to the focus school

This case study examines an independent high school in South Africa that aims to provide high-quality education to learners from low-income communities at an affordable cost [22]. The institution was established in response to concerns about the accessibility of excellent education, the need for inspiration in the South African educational landscape, and the role of a values-driven approach in modern schooling. Its overarching vision is to develop a sustainable and scalable model for delivering top-tier education to underprivileged learners.

Founded in 2018 as an all-boys high school, the institution has since expanded by opening two mixed-sex campuses in 2022 and 2025. In addition to operating these campuses, the school contributes to thought leadership in low-fee education, engaging in policy dialogue and the development of best practices within the broader education sector.

The school's philosophy is rooted in a values-oriented approach, emphasising personal growth, mentorship, and academic excellence. Its emblem symbolises a commitment to fostering a sense of purpose and legacy among learners. Central to this philosophy is the belief that each learner has a unique calling, which they are encouraged to pursue with resolve and diligence. The school's motto reflects this ethos, urging learners to seize opportunities and strive consistently toward their goals. The school follows the national CAPS curriculum, as described in Section 1.1.1.

Despite its values-based identity, the institution is inclusive of learners from all backgrounds and does not require adherence to any particular religion. It promotes an atmosphere of respect, critical thinking, and engagement with societal issues. Opportunities for personal development include structured mentorship programmes, character-building initiatives, and support systems designed to assist learners in managing both academic and personal challenges.

The school operates on a low-fee model, with learners paying a modest annual fee while additional operating costs are subsidised through fundraising. This model helps ensure that high-quality education remains accessible to learners from low-income households. Two primary fundraising strategies support this model: An annual extreme sports event and ongoing financial partnerships.

A flagship fundraising initiative is the annual endurance challenge, which integrates three of Cape Town's most demanding physical feats into a single-day event. Participants must swim 8 km from Robben Island to Bloubergstrand, cycle 109 km in the Cape Town Cycle Tour, and complete the Three Peaks Challenge—a 50 km trail run with 3 000 m of elevation gain—all within 24 hours. Initiated by a supporter of the school, this event has become a powerful symbol of commitment and perseverance. Many learners and teachers participate, recognising the value of raising funds for education. Their involvement fosters a sense of ownership and gratitude for the educational opportunities made available. In 2025, the event raised over R2.3 million [84].

The school relies on the generosity of individuals and organisations that share its vision for accessible, high-quality education. Financial contributions are structured through a learner partnership model, allowing donors to directly support learners. These partnerships assist in covering operational costs and contribute to the institution's long-term sustainability. Fundraising efforts also serve to broaden community engagement and mobilise active support for educational equity. By expanding its donor base, the school maintains a modest fee structure while continuing to deliver a robust learning experience.

Acknowledging that external barriers often hinder academic achievement, the school provides support to ensure learners' basic needs are met. Daily transport is provided to reduce logistical and financial barriers to attendance. All learners receive daily meals to enhance cognitive function and focus. Additional support includes mentorship and emotional guidance to foster academic and personal development. Physical activity is also embedded into the daily schedule, reinforcing general well-being and discipline.

The admissions process is designed to align prospective learners with the school's values-based educational model. Admission is open to learners from diverse backgrounds who demonstrate potential and a willingness to embrace the school's ethos. While academic performance is considered, significant emphasis is placed on character, motivation, and alignment with the institution's mission. Applicants must submit academic records, a motivation letter, and references.

Shortlisted candidates are invited to interviews with school leadership and staff to evaluate their suitability. This process ensures that learners benefit from and contribute to the school environment.

Since its inception, the school has demonstrated that high-quality, independent education can be delivered at low cost. With over 600 learners across its campuses and a dedicated staff of educators, the institution continues to refine and expand its model. Its impact extends beyond enrolled learners, contributing to national discourse on equitable education in South Africa. Through strategic partnerships and innovative practices, the school remains committed to reducing educational disparities and enabling long-term learner success.

5.2 Overview of the data received

During 2025, the focus school's management team administered a survey containing the recommended questions outlined in Section 4.6, along with several additional items. The survey was completed by 271 learners across the school's three campuses. Of these, 11% (*i.e.*, 28 learners) were in Grade 8, 22% (*i.e.*, 60 learners) in Grade 9, 22% (*i.e.*, 60 learners) in Grade 10, 32% (*i.e.*, 87 learners) in Grade 11, and 13% (*i.e.*, 36 learners) in Grade 12. Only three learners reported repeating their previous grade. Regarding academic performance, 58% achieved an average mark of 65% or higher in their 2024 end-of-year report.

Although the school provides daily transport and meals to all learners, 3% reported walking to school, and 1% indicated that they paid additional fees for daily meals. Furthermore, 46% of learners stated that they ate breakfast at home before arriving at school.

Parental involvement was measured along the three dimensions of emotional, physical, and financial support. A total of 58% of learners reported that their father was emotionally involved, defined as engaging in regular conversations and checking on their well-being. Additionally, 66% stated that their father was physically present at key events and offered support when needed. Financially, 77% indicated that their father contributed to household expenses such as food, accommodation, and school fees. Only 58% of learners lived with both a mother and a father figure.

Basic living conditions were generally favourable. A total of 98% of learners had access to clean drinking water, and 84% lived in environments with minimal pollution. Responses regarding government grants suggest that many learners may not have been fully aware of the financial support received by their families. Nevertheless, 94% reported having sufficient food to meet daily needs, 88% noted regular waste removal services, and 97% confirmed access to a flushing toilet.

Perceptions of the school environment were highly positive. A total of 90% of learners believed they received a high-quality education, and 95% stated that their teachers were consistently present. An equal percentage felt their school had a sufficient number of teachers, and all class sizes remained below 35 learners. In terms of infrastructure and resources, 90% reported that their school provided adequate learning materials such as desks, whiteboards, and chairs, and 89% agreed that the buildings were in good condition. Additionally, 93% of learners stated that their teachers had not recently participated in strike action, and a similar proportion reported positive professional relationships among teachers.

Learners also exhibited strong foundational cognitive skills. A total of 91% stated that they could read and comprehend texts independently, 98% could perform basic arithmetic such as calculating change, and 95% were able to write letters. These findings suggest relatively high levels of literacy and numeracy among the learner cohort.

Mental well-being and school motivation were assessed. A total of 83% of learners expressed a positive sentiment toward their school experience, and 91% believed that their education was preparing them for future success. Additionally, 93% reported that someone at school had taught them about their calling and purpose in life, while 75% stated that the school inspired them to think about their future. Almost all learners (99%) agreed that the school promoted good moral values. However, only 60% reported feeling motivated to attend school regularly. Finally, 93% stated that the school provided activities that support their physical health and overall well-being.

5.3 Variate relationship analysis

The theoretical framework of techniques outlined in Section 3.4 is empirically validated in this section through its application to the focus school. To investigate the variate relationships within the school, the dataset is analysed using FA, MRA, and SEM.

5.3.1 Significant factors for the focus school

To gain a comprehensive understanding of the factors influencing learners in the focus school, FA was conducted. EFA was applied to a predefined set of indicators, detailed in Table A.3 in Appendix A. The suitability of the data for EFA was confirmed by an overall KMO-MSA of 0.79, indicating sufficient common variance among the variables to justify the use of FA. Table 5.1 presents the relevant indicators, their corresponding KMO-MSA values, and the factor loadings for the three identified factors. For improved interpretability, the factor loadings were as before scaled by a factor of 100. CFA yielded excellent fit indices (CFI = 0.97, SRMR = 0.046, RMSEA = 0.049).

TABLE 5.1: *KMO-MSA and the rotated factor pattern for case study school characteristics.*

Indicator	KMO-MSA	`Factor 1:` Parental support	`Factor 2:` Mental well-being	`Factor 3:` School resources
`Father's physical support`	0.70	**90***	2	5
`Father's financial support`	0.78	**74***	-5	-6
`Father's emotional support`	0.76	**67***	8	-6
`Co-residence with parents`	0.74	**62***	-10	18
`School satisfaction`	0.81	-1	**76***	31
`Future enthusiasm`	0.78	3	**71***	19
`School motivation`	0.81	-4	**63***	16
`Successful life preparation`	0.85	-1	**47***	28
`Teacher availability`	0.79	8	13	**67***
`Learning materials`	0.82	-4	31	**57***
`School infrastructure`	0.82	-5	25	**48***
`School values promotion`	0.86	6	15	**41***

The three identified factors represent distinct dimensions of learner characteristics at the case study school:

1. `Factor 1` captures variance associated with parental support, clustering indicators related to the emotional, physical, and financial involvement of fathers, as well as the co-residence of learners with their parents.

2. `Factor 2` groups indicators reflecting learners' mental well-being, including positive feelings toward school, excitement about the future, motivation to attend school, and perceived likelihood of life success.

3. `Factor 3` clusters indicators pertaining to school resources, such as the adequacy of teaching staff, availability of learning materials, condition of school infrastructure, and the promotion of positive values within the school environment.

The total variance (R^2) explained by the three factors is 46.29%, indicating that nearly half of the observed variability is accounted for by these latent constructs. This suggests that the factor structure effectively captures the primary dimensions underlying the dataset. The remaining variance may reflect individual differences, measurement error, or other latent variables not captured by the current model. The proportion of variance explained by each factor is presented in Table 5.2.

TABLE 5.2: *Variance explained by the extracted factors for learners in the case study school.*

Factor	R^2
`Factor 1`: Parental support	18.32%
`Factor 2`: Mental well-being	15.58%
`Factor 3`: School resources	12.12%

Table 5.3 presents the Cronbach alpha values for the three extracted factors. These reliability coefficients indicate the internal consistency of the indicators within each factor, demonstrating that the grouped indicators are sufficiently correlated to measure the same underlying construct. The high Cronbach alpha values for `Parental support` (0.81) and `Mental well-being` (0.77) suggest strong internal consistency, while the slightly lower value for `School resources` (0.65) still indicates an acceptable level of reliability.

TABLE 5.3: *Cronbach's alpha values for the extracted factors.*

Factor	Cronbach's alpha
`Factor 1`: Parental support	0.81
`Factor 2`: Mental well-being	0.77
`Factor 3`: School resources	0.65

5.3.2 Factor regression for the focus school

MRA was conducted on the indicators retained through FA, with the inclusion of two additional variables: `Pass grade` and `Pass grade` > 65%. These capture learner attrition and cognitive development through school progression. The `Pass grade` variable reflects whether a learner successfully completed the previous academic year, while `Pass grade` > 65% indicates whether a learner achieved an overall average exceeding 65% in the preceding grade. This threshold was chosen because it aligns with the minimum admission requirements for most university programmes in South Africa. The researchers identified the distinction between learners who qualify for university and those who merely meet minimum progression requirements as a topic of interest for further research [94].

The analysis included fourteen indicators to explore their interrelationships. The resulting regression coefficients are presented in Table 5.4. Statistical significance was defined at $p < 0.05$ (*), with high statistical significance denoted at $p < 0.001$ (**). A colour gradient in each cell visually represents the strength of each regression coefficient, with darker shades indicating stronger associations.

The MRA revealed several significant relationships. The model predicting grade pass demonstrated excellent discriminative performance, with a c-statistic of 0.981. In contrast, the model for grade mark >65% achieved only moderate discrimination, with a c-statistic of 0.655. According to McFadden's pseudo R^2, the grade pass model explains approximately 56.7% of the variation in grade completion, while the grade mark >65% model explains only 7.2%. This discrepancy suggests that the model accounts well for basic academic progression but less effectively for high academic achievement. Additional variables may be required to better understand the factors that drive academic excellence.

School satisfaction positively predicted school motivation ($r = 3.83$, $p < 0.0001$), and the relationship was significantly bidirectional. Learners who were satisfied with their school experience were more likely to be motivated, and motivated learners tended to report greater satisfaction.

The findings illustrate the deep interconnections between various forms of paternal support and their influence on learner well-being. Father physical support is positively associated with co-residence with parents ($r = 2.76$, $p < 0.0001$), suggesting that shared living arrangements facilitated consistent father-child interactions. In turn, co-residence may encourage greater paternal presence. Father physical support also strongly predicts father emotional support ($r = 2.65$, $p < 0.0001$), with a bidirectional association. It is reasonable to conclude that fathers who are physically involved are more likely to provide emotional guidance as well. Collectively, these relationships show the importance of a holistic approach to paternal involvement to contribute to the emotional stability and resilience of learners.

Future enthusiasm positively predicted life preparation ($r = 2.71$, $p < 0.0001$), again in a bidirectional relationship. Learners who felt excited and hopeful about the future tended to perceive themselves as being better prepared for life beyond school, and this confidence may further fuel their engagement with school and personal development.

Finally, teacher availability positively predicted access to learning materials ($r = 2.70$, $p < 0.0001$), with the relationship being significantly bidirectional. When teachers were consistently present and engaged, learners were more likely to have access to essential learning resources. At the same time, well-resourced environments may help retain and motivate teaching staff, suggesting a positive feedback loop between teacher presence and classroom infrastructure.

Teacher availability positively predicted learning materials ($r = 2.70$, $p < 0.0001$), and the relationship is significantly bidirectional. When teachers were consistently available and engaged, learners were more likely to have access to essential learning materials, either through direct provision or facilitated access.

5.3.3 Factor mapping for the focus school

SEM was conducted on the learner data from the case study school. The results are presented in Table 5.5. The outcome of the SEM yielded ideal fit indices (CFI = 0.98, SRMR = 0.049, and RMSEA = 0.034), indicative of a well-fitting model.

TABLE 5.4: *Regression matrix for the indicators based on MRA results for learners in the focus school. Rows represent predictors, columns represent outcomes. Cell shading intensity indicates the magnitude of statistically significant coefficients. ** indicates $p < 0.01$, * indicates $p < 0.05$.*

	Pass grade	Pass grade >65%	Father's emotional support	Father's physical support	Father's financial support	Co-residence with parents	School satisfaction	Future enthusiasm	School motivation	Successful life preparation	Teacher availability	Learning materials	School infrastructure	School value promotion
Pass grade	1.00	15.49	0.53	0.52	0.48	-0.48	1.63	0.70	-2.83	-1.42	3.27	1.43	1.72	-2.21
Pass grade >65%	14.54	1.00	0.14	0.69	0.28	-0.02	-0.06	-0.16	0.87**	0.12	-0.96	-0.40	-0.09	24.81
Father's emotional support	0.94	0.10	1.00	2.61**	2.39**	-0.68	0.41	-0.34	0.51	-0.43	0.09	0.04	-0.86	20.77
Father's physical support	0.52	0.85	2.65**	1.00	2.26**	2.76**	0.83	1.20*	-0.37	-0.15	1.63	-1.04	-0.39	0.65
Father's financial support	0.09	0.04	2.37**	2.26**	1.00	1.58**	-1.41	-0.78	0.46	0.45	-1.39	0.11	-0.35	0.58
Co-residence with parents	-0.48	-0.02	-0.72	2.66**	1.44**	1.00	0.02	-0.10	-0.73	-0.30	1.07	0.26	1.22*	0.12
School satisfaction	1.89	-0.06	0.41	0.14	-0.57	0.02	1.00	2.16**	3.83**	0.84	-0.24	0.93*	1.50*	11.92
Future enthusiasm	-0.79	-0.21	-0.22	1.20*	-0.49	-0.19	1.99**	1.00	1.46**	2.71**	-0.86	1.55*	-0.70	12.24
School motivation	-3.02	0.87**	0.51	-0.37	0.46	-0.73	3.83**	1.41**	1.00	0.29	2.01*	-0.09	1.35*	9.36
Successful life preparation	-1.42	0.31	-0.43	-0.15	0.45	-0.30	0.84	2.41**	0.35	1.00	1.47	-0.04	0.41	17.09
Teacher availability	3.46	-0.96	0.09	1.63	-1.39	1.07	-0.24	-0.35	2.01*	1.47	1.00	2.70**	1.14	6.55
Learning materials	2.66	-0.39	0.04	-1.36	0.07	0.26	0.93*	1.55*	-0.09	-0.04	2.31**	1.00	1.00	20.55
School infrastructure	1.72	-0.09	-0.86	-0.39	-0.35	1.22*	1.70*	-0.70	1.35*	0.41	1.14	1.00	1.00	-5.27
School value promotion	-2.21	24.81	20.77	0.65	0.58	0.12	11.92	12.24	9.36	17.09	6.55	20.55	-5.27	1.00

TABLE 5.5: *Standardised effects of predictors on outcomes including the estimated path coefficients (β), standard errors (SE), and t-values for each relationship of the fourteen indicators.*

Predictor	**Outcome**	β	**SE**	t**-value**
Father physical	Father financial	0.65	0.04	18.38
Father physical	Father emotional	0.60	0.04	15.32
Father physical	Co residence	0.47	0.06	7.91
School satisfaction	Future enthusiasm	0.43	0.05	8.07
School motivation	School satisfaction	0.38	0.05	7.35
Teacher availability	Learning materials	0.34	0.05	6.47
School satisfaction	School infrastructure	0.33	0.05	6.14
Learning materials	School satisfaction	0.30	0.05	6.23
Future enthusiasm	School motivation	0.29	0.06	4.78
Pass grade	School value promotion	0.28	0.06	4.92
School infrastructure	Teacher availability	0.27	0.06	4.83
Teacher availability	Pass grade	0.24	0.06	4.12
School value promotion	Learning materials	0.22	0.05	4.08
School value promotion	Teacher availability	0.18	0.06	3.18
Learning materials	Future enthusiasm	0.17	0.05	3.25
Father physical	Pass grade >65%	0.17	0.06	2.95
Father financial	Co residence	0.17	0.06	2.66
School motivation	Pass grade >65%	0.17	0.06	2.86
Teacher availability	School motivation	0.13	0.06	2.29
Pass grade >65%	Pass grade	0.12	0.06	2.09
School infrastructure	Co residence	0.12	0.05	2.42

The SEM analysis reveals that a father's physical support is strongly linked to his financial support, with learners whose fathers provided financial contributions also being likely to have had physically present fathers ($\beta = 0.65$, $SE = 0.04$, $t = 18.38$). Financial and physical support often go hand in hand, creating a dual support structure. Emotional support from fathers is also strongly connected to physical involvement, with learners who experienced physical support also benefiting from emotional guidance ($\beta = 0.60$, $SE = 0.04$, $t = 15.32$). Additionally, learners who lived with both of their parents tended to have fathers who are consistently physically present ($\beta = 0.47$, $SE = 0.06$, $t = 7.91$), providing a stable home environment.

Learners who were excited about their future often reported feelings about school satisfaction ($\beta = 0.43$, $SE = 0.05$, $t = 8.07$). Future enthusiasm and school satisfaction are closely intertwined, as learners who are content with their school experience tend to carry this optimism forward. Moreover, learners who find satisfaction in their learning environment are more inclined to be motivated to attend school ($\beta = 0.38$, $SE = 0.05$, $t = 7.35$).

Teacher availability is significantly associated with the provision of learning materials. Schools with greater access to educational resources are more likely to have sufficient teaching staff ($\beta = 0.34$, SE = 0.05, t = 6.47). Furthermore, the quality of school infrastructure is positively linked to learner satisfaction, suggesting that learners who perceived their schools to have better physical environments tended to report greater contentment with their educational experience ($\beta = 0.33$, SE = 0.05, t = 6.14). The physical context in which learning takes place thus has a direct influence on learners' emotional engagement. Similarly, the availability of learning materials is strongly associated with school satisfaction ($\beta = 0.30$, SE = 0.05, t = 6.23).

The SEM results largely corroborate the findings derived from MRA, although with some variations in the strength of the observed relationships. The SEM offers a more integrated perspective on the interplay between factors, thereby enabling a more nuanced understanding of the ways in which these variables influence each other within a complex system. The alignment between the two analytical approaches enhances confidence in the robustness of the identified associations. Figure 5.1 presents a network diagramme that illustrates the estimated path coefficients, mapping the relationships between the fourteen indicators and outcomes for the learners within the case study school.

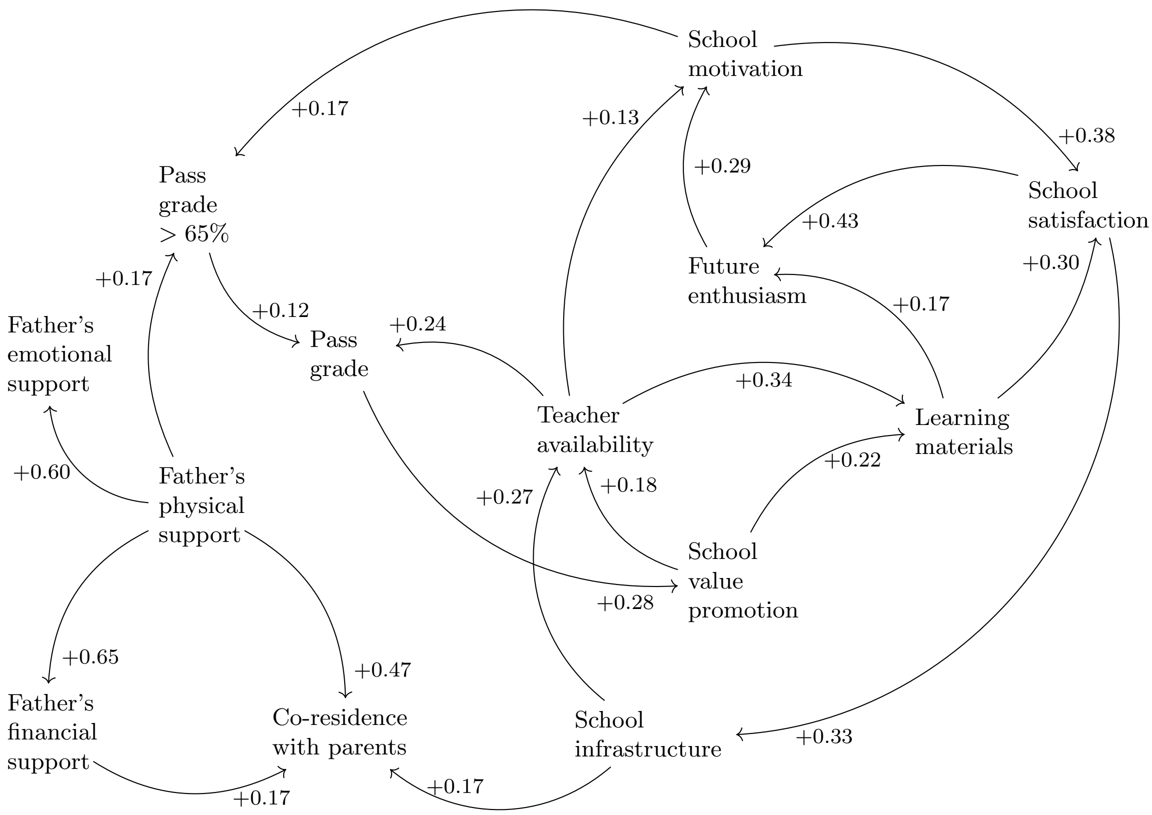

FIGURE 5.1: *Relational mapping of the predictors and outcomes included in the estimated path coefficients as a network diagram of the fourteen indicators for the learners in the case study school.*

5.4 Factor ranking for the focus school

For the feature importance analysis of learner success within the focus school, the dataset comprised 271 observations and 13 predictor variables. As only 1% of learners did not pass, this outcome was excluded from the ML analysis due to significant class imbalance. Consequently, only two outcome categories were considered: learners who passed, and those who passed with an average final mark exceeding 65%. The optimal model parameters, identified through grid search, are summarised in Table 4.2. Model performance was evaluated using precision, recall, and the F1-score. As shown in Table 5.7, the model achieved a precision of 0.67 in predicting learners who attained an average grade exceeding 65%. Precision indicates the reliability of the model's positive classifications. Recall was 0.81 for this group. The F1-score reached 0.74, the results indicating the model's overall effectiveness in predicting academic success among learners.

A confusion matrix, as depicted in Figure 4.2, provides a performance evaluation for classification models. It outlines the model's predictions against the actual outcomes, categorising results into true positives, true negatives, false positives, and false negatives.

TABLE 5.6: *Key gradient boosting classifier parameters for learners in the focus school*

Parameter	Value
`learning_rate`	0.1
`n_estimators`	100
`max_depth`	7
`min_samples_split`	2
`min_samples_leaf`	1
`loss`	log_loss
`subsample`	1.0
`validation_fraction`	0.1

TABLE 5.7: *Classification metrics by class for learners in the focus school*

Class	Precision	Recall	F1-Score
Pass	0.72	0.55	0.63
Pass >65%	0.67	0.81	0.74

The final results showed an overall prediction accuracy of 69.14%. As shown in Figure 5.2, the model correctly classified 55.3% of learners in the 'pass' category and 81.4% in the 'Pass >65%' category. However, 44.7% of the negative cases were incorrectly predicted as positive, and 18.6% of the positive cases were misclassified as negative. These results indicate that the model performs well in identifying high-achieving learners but struggles to accurately classify those in the lower-performing 'pass' group.

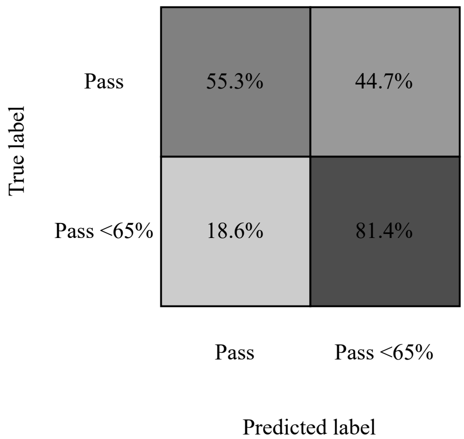

FIGURE 5.2: *Confusion matrix for learners in the focus school.*

Figure 4.3 presents the feature importance derived from the ML model. The most influential predictor was the learner's `Grade`, which reflects their current year level and serves as a contextual anchor for interpreting performance. More notably, however, the next most important features, `Future enthusiasm` and `Father's physical support` highly ranks the powerful role of internal motivation and social support structures in learner achievement.

`Future enthusiasm`, which captures a learner's optimism and excitement about their future, emerged as the second most predictive factor. Hope, aspiration, and future orientation are significant in shaping academic outcomes, especially in schools where material and institutional support may be constrained. Closely following was `Father's physical support`, pointing to the importance of active, hands-on parental involvement. In this context, such support may include helping with schoolwork, attending school meetings, or providing encouragement and supervision—forms of engagement that have a measurable effect on learner success.

Strikingly, features related to school resources and institutional inputs such as `School infrastructure`, `Teacher availability`, and `Learning materials` ranked noticeably lower in importance. This may indicate that, in this particular lower-quintile school, these basic institutional supports are already being provided at a sufficiently consistent level, reducing their variability as predictors of learner success. As a result, the differentiating factors in achievement appear to lie elsewhere, namely in emotional resilience, family involvement, and internal motivation. While continued investment in school infrastructure remains important, targeted efforts to bolster learners' self-belief, future orientation, and parental support may yield more direct benefits in this setting.

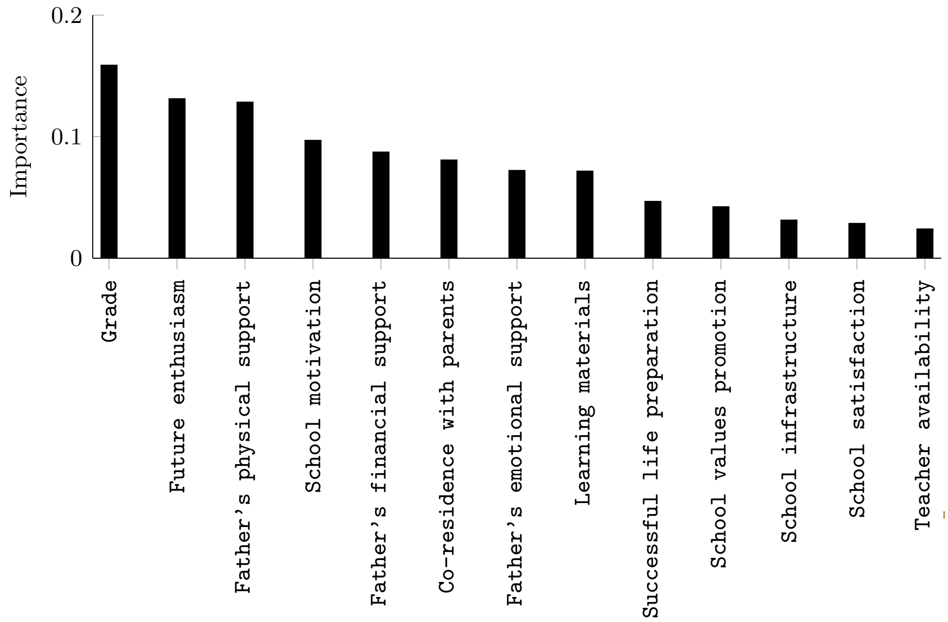

FIGURE 5.3: *Feature importance for learners in the focus school.*

5.5 Chapter 5 summary

This chapter built on the findings presented in Chapters 3 and 4 with the aim of analysing the determinants of learner progression within an independent high school in South Africa, hereafter referred to as the focus school, in alignment with Objective IV. The analysis began with providing contextual information about the selected school. The dataset obtained from the school was summarised. Variate relationships were subsequently examined using FA, MRA, and SEM. Finally, ML techniques were applied to identify the learner attributes most strongly associated with successful academic progression.

CHAPTER 6

Critical synthesis and hierarchical needs decision-support framework development to inform educational interventions

Contents

This chapter critically evaluates the findings from the analyses of the public high school system and the case study school in order to inform the development of a decision-support hierarchical needs framework intended to guide educational interventions for learners from lower socio-economic backgrounds in South Africa in pursuit of Objective V. The chapter begins with a comparative analysis of the results derived from applying the methodological framework to the South African public high school system, as detailed in Section 6.1. This is followed by an examination of the validity of the hypotheses and conclusions through the application of the framework to a case study school, presented in Section 6.2. Finally, Section 6.3 introduces the education hierarchy of needs decision-support framework that emerges from the findings of this study and Section 6.4 briefly discusses the intention for application of the decision-support framework.

6.1 Disparate South African educational systems compared

South Africa's education system continues to grapple with academic decline, socio-economic disparities, and bimodal performance patterns that critically affect youth development and long-term success. Various analysts have worked to uncover the underlying dynamics of the South African education system to inform more effective policy interventions. Slamang [83] employed a systems thinking approach to model the factors influencing academic performance, while Becker [12] investigated learner progression using statistical analysis and relational mapping. Both studies revealed a range of influential socio-economic and institutional determinants. Spaull [92] drew attention to the stark bimodality in educational outcomes, showing two divergent systems stratified by language, school type, and socio-economic status. By integrating the methodologies of Slamang, Becker, and Spaull, this study enables a systemic exploration of the dual schooling pathways within South Africa.

Building on Spaull's analysis of bimodality, this study conducted a comprehensive investigation of lower socio-economic learners (LQ13) across two contrasting educational contexts: Low-income schools (SQ13) and higher-income schools (SQ45) within the public high school system. The study addressed key research questions regarding the characterisation of learners from economically marginalised communities, the interrelationships among influential factors, and the extent to which these factors predict learner progression.

A robust methodological framework was employed, comprising FA, MRA, SEM, and ML techniques. FA enabled the identification of latent constructs describing these learners, while MRA and SEM clarified the complex interrelationships among these constructs. ML techniques further facilitated the ranking of features by predictive importance, offering insights into the most influential factors in learner progression.

FA revealed an identical set of eight latent factors for LQ13 learners in both SQ13 and SQ45 schools. These factors included family structure, learner health, supported retention, food security, sanitation, energy, welfare income, and environmental quality. Together, these constructs comprehensively describe the key dimensions of low-income learners' lived experiences within the South African public high school system. The variance explained by each factor within the two school contexts is summarised in Table 6.1. Although the magnitude of variance explained differs across contexts, the hierarchical order of explanatory strength remains consistent across both systems.

TABLE 6.1: *Variance explained by the extracted factors for LQ13 learners in SQ13 and SQ45 schools respectively.*

Factor	R^2 in SQ13	R^2 in SQ45
`Person F1`: Family structure	22.10%	22.43%
`Person F2`: Learner health	17.60%	20.15%
`Person F3`: Supported retention	17.20%	17.60%
`Hhold F1`: Food security	20.51%	17.49%
`Hhold F2`: Sanitation	14.98%	16.14%
`Hhold F3`: Energy	13.47%	14.81%
`Hhold F4`: Welfare income	13.20%	13.22%
`Hhold F5`: Environment	11.74%	12.22%

It is particularly noteworthy where the largest differences in variance between the two systems occur. For instance, the variance attributed to learner health in the learner's personal profile is 17.60% for LQ13 learners in SQ13 schools, but increases to 20.15% for learners in SQ45 schools.

This indicates that a greater proportion of learner health-related variation is captured in the context of higher-income schools.

There is also a notable difference in the variance explained by food security with 20.51% in SQ13 schools compared to 17.49% in SQ45 schools. Similarly, a difference is evident in sanitation, with 14.98% in SQ13 and 16.14% in SQ45 schools. These results suggest that the available data better capture the nuances of food security among learners in SQ13 schools, while more detailed sanitation-related information is evident for learners in SQ45 schools.

Although learners in both school systems are described using the same eight latent factors, this shared factor structure does not imply uniformity in their lived experiences. Table 3.17 illustrates these differences through learner profiles derived from a "goodness score" metric. Values closer to one represent more advantaged conditions, while values near zero indicate relative disadvantage.

The analysis revealed substantial differences between the two groups, especially in supported retention, welfare income, and sanitation. LQ13 learners in SQ13 schools demonstrate higher supported retention and are more frequently recipients of welfare income, which may be linked to the schools' targeted feeding schemes. A key insight is the discrepancy in welfare income despite similar socio-economic status. This may suggest that learners in SQ13 schools are more likely to apply for government grants, while their counterparts in SQ45 schools may rely on scholarships or other non-state funding mechanisms. Interestingly, LQ13 learners in SQ45 schools report better household sanitation, possibly due to residential proximity to better-resourced schools in more affluent areas. These findings reinforce the importance of context when interpreting shared latent factors across different school settings.

Based on the eight latent factors identified through FA and the key relationships revealed through MRA, SEM was employed to evaluate the interrelationships among these factors within the broader system. The network visualisations of composite factor relationships in SQ13 and SQ45 schools are presented again in Figure 6.1 and Figure 6.2 respectively, for reference convenience. The factors most prominently associated with successful learner progression are highlighted in red.

In particular, the SEM results for LQ13 learners in SQ13 schools reveal five key relationships, as illustrated in Figure 6.1. A closer examination of specific indicator-level relationships, as shown in Figure 3.3, provides a more detailed understanding of the internal dynamics of the system. As discussed in §3.5.3, food security and energy security do not exhibit direct associations with school-related factors and are therefore omitted from the subsequent discussion.

Households located in clean environments are more likely to be both food and energy secure. Learners who receive welfare income are more likely to attend schools with support programmes. This finding supports the earlier hypothesis that, although learners may fall within the same socio-economic classification, those attending SQ13 schools are possibly more actively encouraged to apply for government grants, whereas their counterparts in SQ45 schools. Welfare income is characterised by three indicators: Receipt of a general grant, receipt of a childcare grant, and unearned income. School support programmes are described by indicators including learner retention, participation in school feeding schemes, and walking to school. Among these, the most significant relationship between individual indicators is observed between the receipt of a childcare grant and access to school feeding schemes. This suggests that learners who receive meals at school are also more likely to benefit from a childcare grant.

Traditional family structures are associated with improved access to sanitation. Learners residing with both parents are more likely to have access to flushing toilets and adequate drinking water within their households. Family structure is characterised by the learner living with both parents and the active involvement of both the mother and father in the learner's life. Sanitation

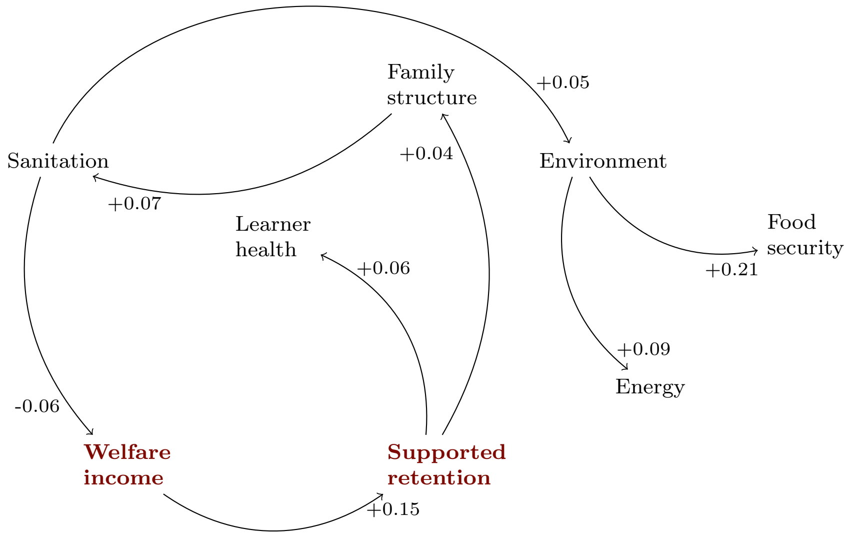

Figure 6.1: *Relational mapping of the predictors and outcomes included in the estimated path coefficients as a network diagramme of the eight composite factors for LQ13 learners in SQ13 schools with highlighted emphasis.*

is defined by the presence of a general or flushing toilet within the household, managed waste collection services, and access to in-house drinking water. Although a strong relationship is evident at the level of the composite factors, no direct relationships emerge between the individual indicators when examined separately.

Attendance at schools that offer support programmes is positively associated with improved learner health. Learners enrolled in such institutions, often characterised by initiatives such as feeding schemes, tend to demonstrate enhanced overall health. Learner health is defined by the presence of good hygiene, memory, self-care, and walking ability. The only direct relationship observed between individual indicators is that between school retention and memory; specifically, learners with good memory are likely remaining within the schooling system.

Building on the results of the SEM analysis for LQ13 learners in SQ45 schools, four key relationships emerge, as depicted in Figure 6.2. A closer analysis of Figure 3.5 highlights the relationships between individual indicators within composite factors, offering deeper insight into the system's internal dynamics. Similar to SQ13 schools, food security, environment, and energy lack direct links to school-related factors and are thus omitted from further discussion.

Learners receiving welfare income are more likely to attend schools offering support programmes, reaffirming the earlier hypothesis linking school-based support to welfare access. Several indicator relationships appear within the welfare income and supported retention composite factors. The strongest relationship is observed between receipt of the child support grant and school retention, indicating that learners attending school are likely receiving this specific grant. Conversely, a counter-intuitive relationship exists between general grant receipt and school retention, whereby learners attending school are less likely to be receiving a non-specific grant.

Households receiving welfare income are more likely to experience food insecurity. While this may initially appear counter-intuitive, the relationship suggests that learners from households

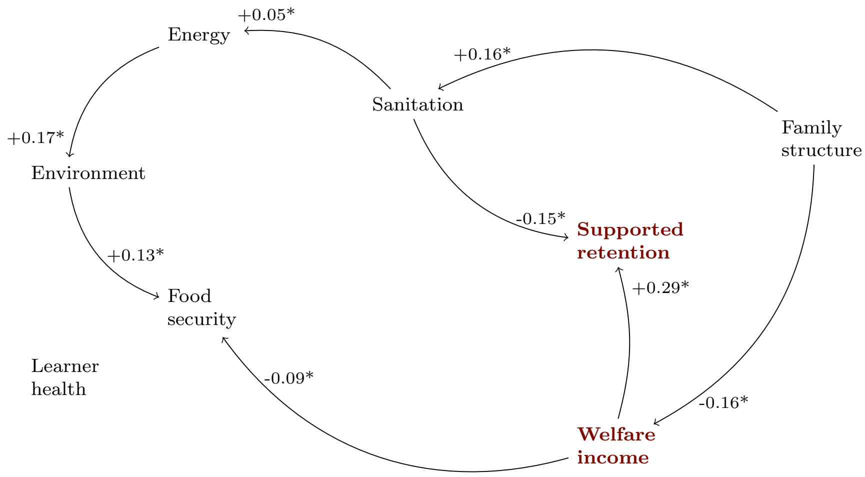

Figure 6.2: *Relational mapping of the predictors and outcomes included in the estimated path coefficients as a network diagrammeme of the eight composite factors for LQ13 learners in SQ45 schools with highlighted emphasis.*

with lower welfare and higher earned income tend to be more food secure, reflecting greater overall household stability.

Traditional families where both parents are present, are less likely to receive welfare income, consistent with prior findings that dual-parent households generally exhibit lower welfare dependency. The strongest indicator relationship appears between co-residence with both parents and receiving unearned income, suggesting that learners in households with less unearned income are more likely to reside with both parents. A less numerically significant, but still noteworthy relationship exists between maternal involvement and receiving a childcare grant. Learners receiving a childcare grant are more likely to have mothers actively involved in their lives, suggesting that maternal participation plays a significant role in securing the specific grant.

Learners from households with good sanitation are less likely to attend schools with support programmes. Since such programmes are typically offered by schools in lower-income communities, and learners often attend schools near their homes, this suggests that those in more affluent areas (characterised by better sanitation infrastructure) tend to attend schools with fewer support interventions, reflecting a higher socio-economic status. The most significant indicator relationship is observed between school retention and residing in a household with in-house drinking water, indicating that learners attending school are likely to also have access to in-house drinking water.

The results yielded through MRA and SEM revealed that welfare income, supported retention, sanitation, and family structure are prominent factors influencing learners both within SQ13 and SQ45 schools. These factors play a significant role in shaping the learner success within the school system.

The application of ML techniques revealed the following feature importance. For lower socio-economic learners attending SQ13 schools, the most influential features contributing to their academic success include regular school attendance, age-appropriate grade placement, receipt of government grants, and access to school meals. Similarly, for lower socio-economic learners in SQ45 schools, the same four features emerge, albeit in a slightly different order of importance.

These features are school attendance, school meals, appropriate age-for-grade placement, and grant receipt.

For learners in lower socio-economic schools, consistent attendance and age-grade alignment are particularly critical, reflecting persistent challenges in the South African education system, such as grade repetition and over-age enrolment. These issues are often linked to delayed school entry, repeated grades, or disruptions in schooling due to socio-economic hardships. The comparatively lower ranking of school meals in this context may indicate the presence of community-level support structures that partially mitigate food insecurity. Conversely, for lower socio-economic learners attending higher socio-economic schools, school meals assume greater importance, possibly reflecting heightened food insecurity relative to their more affluent peers.

Reflecting on the research questions, the findings indicate that the same eight factors characterise learners from lower SES backgrounds across two distinct school systems. However, the initial factor scores and relationships vary between the systems. Learner progression for both cohorts is significantly influenced by attending school and their age-to-grade delta. In addition, the progression of LQ13 learners in SQ13 schools is more sensitive to welfare income, while the progression of LQ13 learners in SQ45 schools is more responsive to the availability of school feeding programmes.

For a low-income learner in South Africa to successfully progress through the education system and attain their NSC at the end of their matric year, ensuring consistent school attendance is of utmost importance. Early intervention to maintain the learner's age-grade alignment is also crucial. Specifically for low-income learners, securing welfare income is vital, and support for the application process could aid learners and their families. Additionally, food security is essential, with access to school meals being particularly significant.

6.2 Case study reflection

An analysis of the public high school system in South Africa suggests that, for a learner from a low-income background to succeed academically, several basic conditions must be met. These include regular school attendance, age-appropriate grade placement, access to daily meals, and the availability of welfare income opportunities. This hypothesis is examined through a case study of an independent high school in South Africa.

The focus school provides high-quality education to learners from low-income communities at a minimal cost. Established in 2018, it follows the CAPS curriculum, thereby aligning with the national standards of South African public education. The school maintains a values-based approach that emphasises personal development and academic achievement. The school operates on a low-fee model relying significantly on external funding obtained through fundraising initiatives and financial partnerships. To reduce barriers to academic success, it offers daily meals and transportation to all learners. In addition, the school adopts a holistic model that includes mentorship, emotional support, and physical activity as integral components of the daily programme.

Given that the focus school, informed by research on the South African public school system (particularly concerning low-income learners) has deliberately addressed the key conditions identified in the earlier analysis (namely, promoting attendance through mentorship, ensuring age-appropriate grade placement, and providing meals), this alignment offers an initial point of confirmation that the proposed hypothesis is well-founded. The subsequent analysis of learner data from the focus school therefore serves as a means to further evaluate the validity of this hypothesis.

In 2025, the management team of the focus school administered a survey to learners, incorporating questions recommended in §4.6 along with additional items tailored to the school's context. The data comprised 271 learner responses. Through the application of FA, three latent constructs emerged: Parental support, learners' mental well-being, and the availability of school resources. At first glance, these findings may suggest that learners from low SES backgrounds attending the focus school differ significantly from those in public schools. However, it is important to recognise that FA identifies variation among variables. Core variables such as regular school attendance, age-appropriate grade placement, school meals, and access to welfare income support do not appear prominently in the analysis—not because they are unimportant, but because all learners at the focus school already possess these attributes. The lack of variance in these foundational factors implies that the basic conditions for learning are uniformly met.

Considering the three latent factors identified through the application of FA, and the key relationships extracted in MRA, SEM was employed to evaluate the interrelationships between the indicators that comprise the factors. Two additional indicators, `Pass grade` and `Pass grade` $>$ `65%`, assess school progression via learner attrition and cognitive performance. In the subsequent discussion of the SEM results, relationships between indicators that originate from the same composite factors are omitted.

Several strong associations were identified between indicators within the composite factors of school resources and mental well-being. Specifically, learners in school environments with adequate infrastructure were more likely to report higher levels of school satisfaction. Similarly, learners who expressed satisfaction with their education were also more likely to have access to sufficient and high-quality learning materials, such as desks, textbooks, and writing tools. Furthermore, learners with access to adequate learning resources and regularly available teachers were more likely to attend a school that promotes a positive and values-driven environment.

With regards to a learner passing their respective grade, the following relationships appeared. Learners who had motivation to attend school are more likely to pass their grade. Learners who passed their grade are likely to have had a teacher who was available and willing to help them. A learner who pass their grade with an average mark of more than 65% most likely had a father who was physically involved in their life. These relationships reveal that for a learner who passes their grade, their mental well-being and a good support structure are of importance.

Building on the findings from the FA, MRA, and SEM, ML was employed to assess which indicators are most important in predicting a learner's pristine progression through the focus school. In this context, pristine progression refers to a learner achieving an average mark exceeding 65%. The most important predictor identified was the learner's current grade, suggesting that academic success is influenced by the stage of schooling. Following this, the learner's enthusiasm for the future, the father's physical involvement, and the learner's motivation to attend school emerged as pivotal variables. These findings reinforce that a learner's progression within the focus school is strongly associated with their mental well-being and support structures. The network visualisation of indicator relationships in the focus school with highlighted emphasis on the variables which emerge as prominent variables in successful learner progression is presented in Figure 6.3.

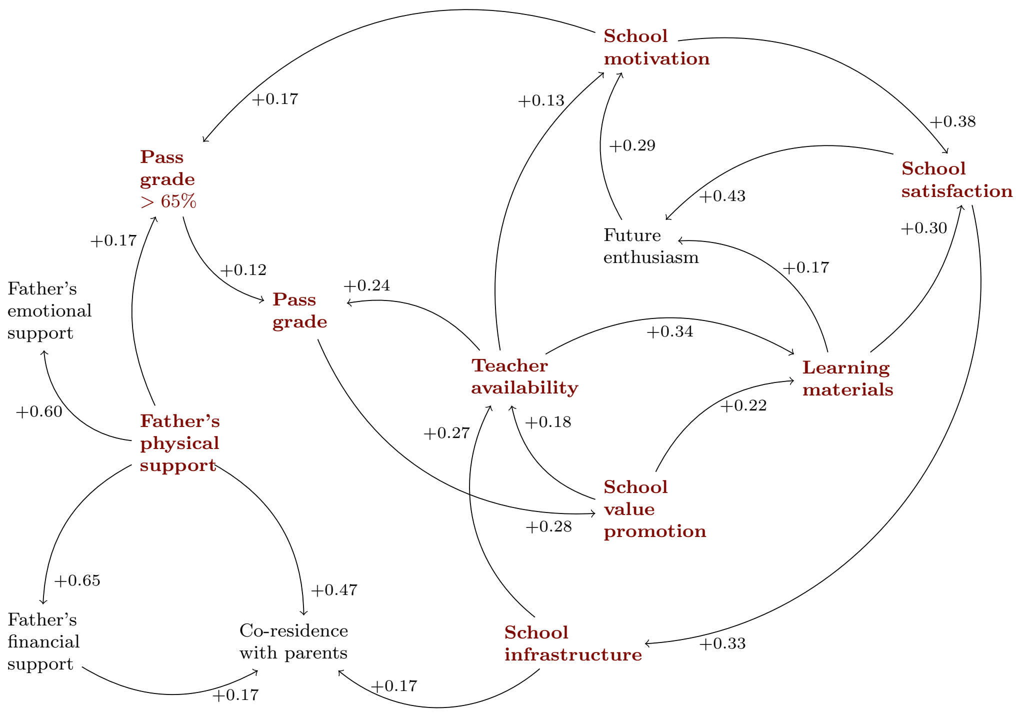

FIGURE 6.3: *Relational mapping of the predictors and outcomes included in the estimated path coefficients as a network diagrammeme of the fourteen indicators for the learners in the focus school with highlighted emphasis.*

6.3 Emergent education hierarchy of needs

The analysis of low-income learners within the South African public education system identified regular school attendance, age-appropriate grade placement, access to meals during the school day, and availability of welfare income as key factors contributing to successful school completion. In contrast, the analysis of low-income learners within the focus school revealed that a father's physical involvement, a learner's motivation to attend school, and their optimism about the future are significant variables influencing their academic success.

To distinguish between the foundational and higher-level needs within a learner's educational trajectory, this study draws on Maslow's Hierarchy of Needs [61]. This framework, visually depicted in Figure 6.4, categorises human needs into five levels. Level 1 represents the most basic physiological needs essential for survival, such as access to food and water. Level 2 encompasses safety needs, including health, security, and protection from harm. Level 3 addresses love and belonging needs, which involve feeling accepted and forming meaningful relationships. Level 4 pertains to esteem needs, including self-respect, recognition, and mental well-being. Finally, Level 5 represents self-actualisation—the fulfilment of personal potential and the pursuit of growth and purpose.

The findings suggest that learners at the focus school have progressed beyond the stage of securing their basic educational necessities. Core need such as regular school attendance, age-appropriate grade placement, access to meals during the school day, and opportunities for welfare income, are consistently met. As a result, these learners are able to move upward within the hierarchy. The analysis reveals the importance of addressing higher-order developmental needs,

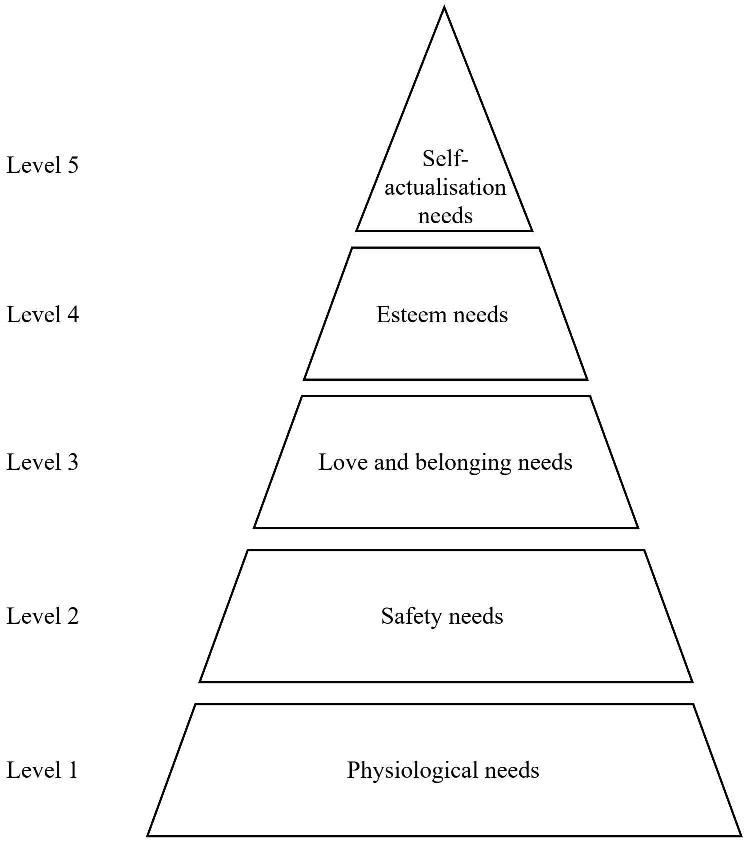

FIGURE 6.4: *Maslow's Hierarchy of Needs [61]*

including emotional well-being, supportive interpersonal relationships, and access to resources that facilitate academic self-actualisation. These findings validate the effectiveness of the school's holistic and values-based approach and support the hypothesis proposed earlier in this study. The evidence indicates that, for a low-income learner, meeting foundational educational needs is the essential first step. Once these are in place, attention must shift toward strengthening support structures and promoting positive mental well-being to facilitate further academic and personal development.

This analysis informs the development of an emergent education hierarchy of needs model for low-income learners in South Africa, as illustrated in Figure 6.5. The hierarchical structure captures the essential educational needs of learners from low socio-economic backgrounds, illustrating how progression through each level supports their holistic development and educational success. As the needs at each level are met, the learner is increasingly equipped to succeed within the schooling system.

At Level 1, foundational needs include regular school attendance and age-appropriate grade placement. Learners who are chronically absent or significantly over-age for their grade are less likely to succeed academically. Level 2 emphasises the importance of receiving meals during the school day. For many learners from disadvantaged backgrounds, this may be the only substantial meal they receive daily. A nourished learner is more likely to concentrate and engage effectively in the classroom.

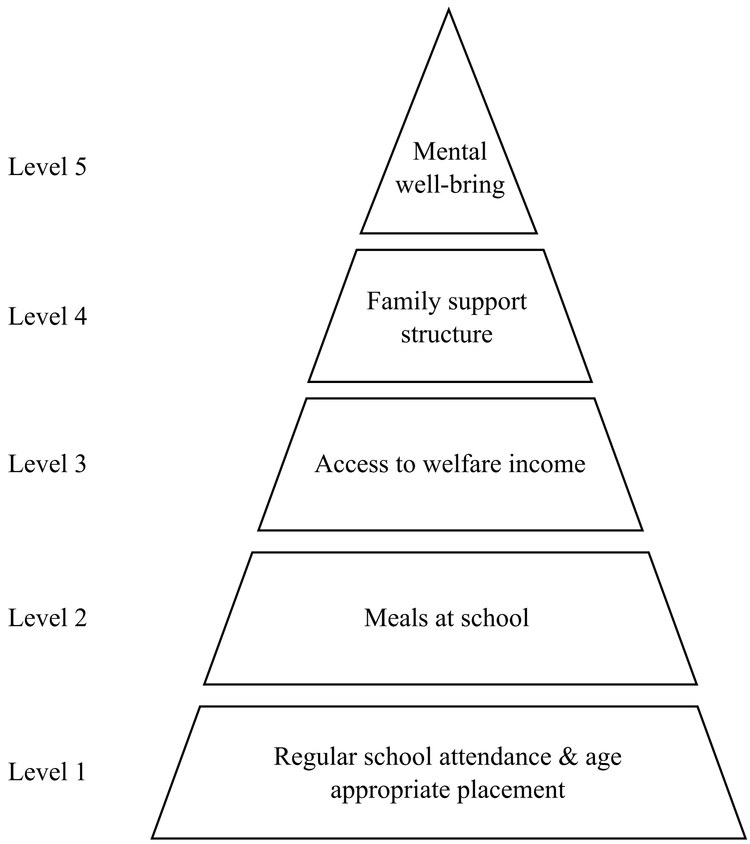

Figure 6.5: *Education hierarchy of needs for low-income learners within South Africa.*

Level 3 concerns access to social welfare, specifically the child support grant. Learners from households receiving this form of financial assistance are generally better positioned to pursue their education. Level 4 addresses the role of family support, particularly the physical involvement of the father in the learner's life, which correlates strongly with educational attainment. Finally, Level 5 focuses on the learner's mental well-being. At this stage, the learner becomes self-motivated, attends school willingly, and holds a positive outlook towards their future.

Importantly, as with Maslow's Hierarchy of Needs, progression must follow a bottom-up approach. For example, a learner with access to welfare income (Level 3) but who does not attend school regularly (Level 1) is unlikely to succeed. Similarly, strong mental well-being (Level 5) cannot compensate for a lack of basic nutrition (Level 2). The structure therefore affirms the cumulative and interdependent nature of these educational needs.

6.4 Application of decision-support framework

The proposed hierarchical framework is not intended as a policy recommendation in itself, but rather as a decision-support tool designed to assist policy-makers in evaluating and improving interventions. The framework aims to enable policy-makers to assess where learners within a given school or context fall within the hierarchy of educational needs. This insight can inform a more targeted and effective response to learners' circumstances. The hierarchical framework provides a useful lens through which to evaluate the appropriateness and sequencing of interventions. Ultimately, this decision-support framework enables policy-makers to assess current learner conditions against the structured model and to make necessary adjustments to policies and interventions, ensuring that resources and support are aligned with learners' most immediate needs.

6.5 Chapter 6 summary

This chapter critically evaluated the findings from the analyses of the public high school system and the case study school to inform the development of a decision-support hierarchical needs framework intended to guide educational interventions for learners from lower socio-economic backgrounds in South Africa, in pursuit of Objective V. The chapter began with a comparative analysis of the results derived from applying the methodological framework to the South African public high school system. This was followed by an examination of the validity of the hypotheses and conclusions through the application of the framework to a case study school. Finally, the education hierarchy of needs decision-support framework that emerged from the findings of this study was introduced, and the intended application of the decision-support framework was discussed.

CHAPTER 7

Conclusion

Contents

This chapter serves to consolidate the findings of the study, critically reflect on its limitations, and propose directions for future research within this domain, in alignment with Objective VI. Section 7.1 provides a summary of the preceding chapters. The project appraisal is presented in Section 7.2, highlighting key findings and the author's concluding reflections. Section 7.3 addresses the limitations of the study and Section 7.4 proposes potential avenues for future development. The chapter concludes with the author's final reflections in Section 7.5.

7.1 Summary

Chapter 1 introduced South Africa's education system and research focus. Post-apartheid South Africa faces significant socioeconomic challenges, including 31.9% unemployment and 34.2% youth neither employed nor in education. The education system includes public schools divided into quintiles by SES, with Quintiles 1-3 receiving more government funding as no-fee schools. Educational outcomes are concerning - international assessments show poor performance in reading and math, with a "bimodal" system where wealthier students significantly outperform poorer ones. The research examines factors affecting lower socioeconomic learners in different school environments and how these factors influence grade progression.

Chapter 2 reviewed methodologies for modeling education systems and positions the current research within the STEP group's work. The research builds on previous STEP work: Venter's PSM found home support most influenced primary learners' performance; Slamang's HSM identified classroom interventions as most important for secondary education; and Becker identified nine critical factors affecting learner progression including family structure and socioeconomic factors. The study addresses Spaull's observation of a "bimodal" South African education system where outcomes are divided along socioeconomic lines, seeking to understand how these dual educational realities function structurally to inform more inclusive educational strategies.

Chapter 3 explored the relationship between factors affecting South African high school learners, specifically comparing those from low socioeconomic quintiles attending different school types. The research used FA, MRA, and SEM on data from the GHS. Eight composite factors were identified: family structure, learner health, supported retention, food security, welfare income, sanitation, energy, and environment. The analysis revealed significant relationships between these factors, such as the strong positive correlation between welfare income and supported retention, and between family structure and sanitation. Those in lower quintile schools received more institutional support and welfare income, while those in higher quintile schools had better home sanitation facilities. The research created "goodness scores" to measure learner advantage levels, finding that supported retention, welfare income, and sanitation showed the greatest disparities between the two school systems, despite serving learners from similar socioeconomic backgrounds.

Chapter 4 examined feature importance in predicting grade progression among South African high school learners from low socioeconomic backgrounds (LQ13) attending different school types. Using gradient boosting ML, the study identified key factors affecting academic success. For both school types, learner retention emerged as the most important factor. In lower quintile schools (SQ13), age appropriateness and child support grants followed in importance, while in higher quintile schools (SQ45), school meals and age were secondary factors.

Chapter 5 applied the methodological framework developed in Chapters 3 and 4 to an independent South African high school, referred to as the focus school. Founded in 2018, the school follows the national CAPS curriculum and adopts a values-driven approach that emphasises personal growth, mentorship, and academic excellence. The research used FA, MRA, and SEM on data from the focus school. Three latent factors were identified: parental support, mental wellbeing, and school resources. The analysis revealed significant relationships between a father's physical involvement, a learner's motivation to attend school, future optimism, and successful academic progression.

Chapter 6 facilitated a ciritical synthesis of the preceding findings. It commenced by evaluating the results derived from the analysis of low-income learners within the South African public high school system, where four critical factors emerged as key influencers of academic success: regular school attendance, age-appropriate grade placement, access to meals during the school day, and receipt of welfare income. In the subsequent case study of an independent school, it was found that once these foundational needs were met, higher-order factors gained significance. Namely, paternal involvement, motivation to attend school, and optimism about the future. These insights informed the development of an educational hierarchy of needs decision-support framework, which emphasises a bottom-up progression model. This decision-support framework asserts that higher-level needs cannot compensate for deficiencies at foundational levels.

7.2 Appraisal of project contributions

This project sought to address the academic decline, socio-economic disparities, and bimodal performance that undermine youth success within South Africa's education system. Building on prior research conducted by STEP researchers, notably Slamang and Becker, who identified a range of socio-economic and institutional determinants. Through the incorporation of Spaull's findings on the stark bimodality dividing the system by language, school type, and socio-economic status, this study considered South Africa's dual education structure. The research aimed to identify the factors characterising public high school learners from lower socio-economic communities across two distinct schooling contexts, to examine how these factors are interrelated within and across these settings, and to assess the extent to which they predict learner progression.

This study introduces a methodological framework that integrates FA, MRA, SEM, and ML to analyse South Africa's bimodal education system. Rather than applying these techniques in isolation, their combined use enabled a comprehensive investigation into the factors influencing learner progression across diverse socioeconomic contexts. The approach identified latent constructs, uncovered complex interrelationships, and assessed the predictive importance of key variables linked to academic success. This integrated analytical pipeline offers a replicable model for future research in varied educational environments.

The research demonstrated that identical factors function differently in SQ13 schools compared to SQ45 schools, despite learners sharing similar socioeconomic backgrounds. Although both systems are shaped by the same eight factors (*i.e.*, family structure, learner health, supported retention, food security, welfare income, sanitation, energy, and environment) their interrelationships diverge. Learners in lower-quintile schools benefit more from institutional support yet face poorer household conditions, particularly regarding sanitation. Conversely, learners in higher-quintile schools reap the rewards from improved home environments but receive less institutional assistance. These findings challenge the efficacy of universal education policies.

The case study, which builds upon the findings derived from the analysis of the public school system, provides critical validation for the hypotheses formulated from the state-level data. By analysing a school that actively addresses foundational needs (*i.e.*, providing meals, transport, and ensuring regular attendance), the research demonstrated that, once these basic conditions are satisfied, learner progression is increasingly shaped by higher-order factors (*i.e.*, paternal involvement, future optimism, and motivation to attend school).

The emergent educational hierarchy of needs decision-support framework offers a critical theoretical advancement in understanding educational progression within low-income communities. Inspired by Maslow's structure, it outlines five ascending levels for low-income South African learners. The model begins with school attendance and grade appropriateness, followed by school meals and welfare access, and culminating in family support and mental well-being. This evidence-based framework prioritises interventions by emphasising that sustainable educational advancement requires the foundational layers to be firmly in place. It challenges intervention strategies that overlook these prerequisites, reinforcing the importance of a bottom-up, sequential approach to supporting learners' academic development.

7.3 Limitations

Several constraints emerged during the research process that could not be addressed within the scope of this project.

Limitation I Data source constraints

> This study draws primarily on the 2022 GHS, which includes fewer indicators of educational quality than earlier iterations. The absence of detailed classroom environment variables introduces a potential bias toward household-related factors. Additionally, the use of single-year, cross-sectional data limits the analysis to a snapshot view, restricting variate relationships and the examination of learner progression over time. The GHS is based on self-reported data, which may introduce reporting bias, particularly in educational and socio-economic indicators, potentially affecting the reliability of key variables in the analysis.

Limitation II Case study generalisability

Validating findings through a single case study school, which follows a distinct educational model, may limit the generalisability of results to the wider South African education system, especially to public schools facing varied resource constraints and contextual challenges.

7.4 Future work

The project results in several potential research directions that could be further explored to enhance and extend the contributions realised in this project.

Proposal I Agent-based simulation model

Future research could benefit from developing an agent-based simulation model of the school system to simulate individual learner trajectories, explore dynamic interactions between influencing factors, and test policy interventions under varying scenarios for more comprehensive, data-driven educational planning.

Proposal II Longitudinal study

Repeating the methodological framework across multiple years of GHS data would enhance the research by revealing trends and shifts in the educational landscape, offering insights into how the needs and circumstances of South Africa's low-income learners evolve over time. Such a study could assist in transforming the static relational maps developed by SEM into dynamic CLDs for future SD simulation modelling analyses.

Proposal III Hybrid system dynamics and agent-based simulation model

Incorporating the longitudinal study with an agent-based simulation model would enable the development of a hybrid system dynamics model. This model could simulate learner progression and track the long-term impact of interventions, offering a dynamic view of how changes in policy or support structures influence educational outcomes. Such a model would provide more practical, evidence-based insights to inform interventions aimed at improving the educational trajectory of low-income learners in South Africa over time.

Proposal III Costing considerations for real-world implementation

While this study does not include a costing analysis of applying the proposed decision-support model, such considerations fall outside the scope of the present research. However, it is important to acknowledge that financial feasibility cannot be disregarded in real-world implementation. This is a theoretical study grounded in empirical data, and future research should address the economic implications associated with operationalising the framework in practice.

7.5 Final thoughts

The unjust nature of South Africa's education system cannot be ignored or dismissed. Instead, it must be acknowledged as a necessary step towards meaningful progress. The two disparate schooling systems must be treated distinctly. The deep-rooted inequities within the current system are not accidental but are, in large part, the enduring legacy of apartheid-era policies compounded by decades of mismanagement and institutional neglect, which systematically disadvantaged the majority of the population. Addressing these inequities requires a concerted effort to elevate the experiences of learners from lower-income, predominantly African communities to those of their more affluent counterparts.

Although the decision-support hierarchical model was developed within the specific context of South Africa, the underlying themes are broadly applicable to a range of global educational settings. In this instance, the model focuses on the progression of low-income learners through the school system, from a state of survival to one of flourishing. The core principle that basic needs must be met before higher-order developmental and academic goals can be achieved, is relevant across diverse contexts.

The decision-support framework is particularly pertinent in other developing or third-world countries where learners often face systemic poverty, limited access to resources, and inadequate school infrastructure. It is equally applicable to learners from marginalised or historically disadvantaged communities in more developed countries, including ethnic minorities, refugees, or learners experiencing socio-economic hardship. These groups, like their South African counterparts, often face significant barriers to educational attainment due to unmet foundational needs such as safety, nutrition, stable housing, and consistent school attendance.

By recognising and responding to these foundational needs, educational systems worldwide can create more inclusive and equitable learning environments. Applying this model internationally provides a practical lens through which policymakers and educators can tailor interventions for vulnerable learners, enabling them not only to participate in education but to thrive within it. In doing so, the model offers a valuable contribution to addressing educational inequality on a global scale, particularly for minority and disadvantaged populations.

Maslow's hierarchy of needs has frequently been employed within educational contexts to explain the notion that learners must have their basic physiological and safety needs met (*i.e.*, food, shelter, and personal security) prior to engaging effectively in higher-order learning. Numerous scholars and practitioners have drawn upon this theoretical framework to articulate the layered nature of learner needs, emphasising the importance of belonging, esteem, and self-actualisation in supporting educational outcomes ([18]; [39]; [71]). However, the application of Maslow's theory within education has remained largely conceptual and generalised.

This study presented a significant advancement by moving beyond theoretical discourse to develop an empirically validated and contextually specific hierarchy of educational needs, designed explicitly for low-income learners in developing countries. This contribution is novel in several respects. It adapts psychological hierarchy theory to specific educational and socio-economic contexts. It is grounded in empirical evidence generated through rigorous analytical techniques, including FA, MRA, SEM, and machine learning. Finally, it addresses a clear gap in existing decision-support education policy frameworks relevant to developing contexts.

The proposed hierarchy offers a sequenced, evidence-based decision-support framework for policy-making. This study transformed a long-standing theoretical model into a practical policy tool. In doing so, it provides policy-makers with structured guidance for prioritising educational interventions in resource-constrained settings, thereby contributing meaningfully to efforts aimed at reducing educational inequalities within developing countries.

Bibliography

[1] 2030 Reading Panel, 2023, *2030 reading panel*, [Online], [Cited February 6th, 2024], Available from www.readingpanel.co.za/.

[2] Aalen O, 1989, *A linear regression model for the analysis of life times*, Statistics in medicine, **8(8)**, pp. 907–925.

[3] Alzubi J, Nayyar A and Kumar A, 2018, *Machine learning from theory to algorithms: An overview*, Journal of Physics: Conference Series, **1142**, p. 012012.

[4] Ammara U, Qudrat-Ullah H, Al-Fuqaha A and Qadir J, 2021, *Using the lens of systems thinking to model education during and beyond covid-19*, Proceedings of the International Wireless Communications and Mobile Computing (IWCMC), Harbin, pp. 2056–2061.

[5] Amra I and Maghari A, 2017, *Students performance prediction using knn and naïve bayesian*, Proceedings of the 5th International conference on information technology (ICIT), Amman, pp. 909–913.

[6] Anderson J and Gerbing D, 1988, *Structural equation modeling in practice: A review and recommended two-step approach.*, Psychological bulletin, **103(3)**, p. 411.

[7] APA Dictionary of Pyschology, 2023, *Socioeconomic status (ses)*, [Online], [Cited January 31st, 2024], Available from www.dictionary.apa.org/socioeconomic-status.

[8] Arafiyah R, Santoso H and Hasibuan Z, 2022, *Monitoring learners' performance by modeling learning progress using machine learning*, Journal of Engineering Science and Technology, **17**, pp. 30–39.

[9] Astakhova K, Korobeev A, Prokhorova V, Kolupaev A, Vorotnoy M and Kucheryavaya E, 2016, *The role of education in economic and social development of the country*, International Review of Management and Marketing, **6(1S)**.

[10] Bagozzi R, 1977, *Structural equation models in experimental research*, Journal of Marketing Research, **14(2)**, pp. 209–226.

[11] Bartlett M, 1954, *A note on the multiplying factors for various chi square approximation*, Journal of the Royal Statistical Society. Series B (Methodological), **16**, pp. 296–8.

[12] Becker E, 2024, *Statistiese kartering van faktorverwantskappe in die suid-afrikaanse openbare hoërskoolstelsel*, Master of Commerce Thesis, Stellenbosch University, Stellenbosch.

[13] Bentéjac C, Csörgő A and Martínez-Muñoz G, 2021, *A comparative analysis of gradient boosting algorithms*, Artificial Intelligence Review, **54**, pp. 1937–1967.

[14] Berger D, 2003, *Introduction to multiple regression*, USA: Claremont Graduate University.

[15] Berger D, 2004, *Using regression analysis*, Handbook of practical program evaluation, **2**, pp. 479–505.

[16] Borsboom D, Deserno M, Rhemtulla M, Epskamp S, Fried E, McNally R, Robinaugh D, Perugini M, Dalege J, Costantini G *et al.*, 2021, *Network analysis of multivariate data in psychological science*, Nature Reviews Methods Primers, **1(1)**, p. 58.

[17] Braun V and Clarke V, 2024, *Thematic analysis*, pp. 7187–7193 in *Encyclopedia of quality of life and well-being research*, pp. 7187–7193. Springer.

[18] Brooks and Kirk, 2024, *Maslow's hierarchy of needs in education*, [Online], [Cited July 8th, 2025], Available from https://brooksandkirk.co.uk/understanding-maslows-hierarchy-of-needs-in-education/.

[19] Byrne B, 1998, *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*, Lawrence Erlbaum, Mahwah (NJ).

[20] Byrne B, 2009, *Structural equation modeling with AMOS: Basic concepts, applications, and programming*, 2nd Edition, Lawrence Erlbaum, Mahwah (NJ).

[21] Cabrera A, 1994, *Logistic regression analysis in higher education: An applied perspective*, Higher education: Handbook of theory and research, **10**, pp. 225–256.

[22] Calling Education, 2025, *Calling education official website*, [Online], [Cited April 2nd, 2025], Available from https://www.callingeducation.org.za/.

[23] CAPS 123, 2023, *South african school quintiles: Understanding the system*, [Online], [Cited February 2nd, 2024], Available from www.caps123.co.za/south-african-school-quintiles-understanding-the-system/#:~:text=South%20African%20schools%20are%20ranked,school%20receives%20from%20the%20government.

[24] CAPS 123, 2023, *Understanding school fees and quintiles in south african public schools*, [Online], [Cited February 2nd, 2024], Available from www.caps123.co.za/south-african-school-quintiles-understanding-the-system/#:~:text=South%20African%20schools%20are%20ranked,school%20receives%20from%20the%20government.

[25] Cattell R, 1966, *The scree test for the number of factors*, Multivariate behavioral research, **1(2)**, pp. 245–276.

[26] Chawla N, Bowyer K, Hall L and Kegelmeyer W, 2002, *Smote: synthetic minority over-sampling technique*, Journal of artificial intelligence research, **16**, pp. 321–357.

[27] Costanza F, 2022, *Covid-related educational policies in action: a system dynamics view*, International Journal of Public Sector Management, **35(4)**, pp. 480–512, Available from https://doi.org/10.1108/IJPSM-07-2021-0187.

[28] Cowling N, 2022, *Number of students in south africa in 2022, by sector*, [Online], [Cited February 26th, 2024], Available from www.statista.com/statistics/1322778/number-of-students-in-south-africa-by-sector/#:~:text=The%20total%20number%20of%20students,in%20the%20independent%20schooling%20sector.

[29] Cowling N, 2023, *Household disposable income per capita in south africa from 2004 to 2022*, [Online], [Cited September 24th, 2024], Available from https://www.statista.com/statistics/874035/household-disposable-income-in-south-africa/#:~:text=In%202022%2C%20South%20African%20households,(around%202%2C725%20U.S.%20dollars).

[30] Cronbach L, 1951, *Coefficient alpha and the internal structure of tests*, psychometrika, **16(3)**, pp. 297–334.

[31] Daoud J, 2017, *Multicollinearity and regression analysis*, Proceedings of the 1st Journal of Physics: Conference Series, Volume 949, Kuala Lumpur.

[32] DataFirst, 2020, *National schools effectiveness study 2007-2009*, [Online], [Cited March 4th, 2024], Available from www.datafirst.uct.ac.za/dataportal/index.php/catalog/694/variable/F3/V3262?name=q92_enggr5_2008.

[33] DataFirst, 2023, *General household survey 2021*, [Online], [Cited February 9th, 2024], Available from www.datafirst.uct.ac.za/dataportal/index.php/catalog/905.

[34] DataFirst, 2024, *General household survey 2022*, [Online], [Cited July 15th, 2024], Available from https://datafirst.uct.ac.za/dataportal/index.php/catalog/945.

[35] Department of Basic Education, 2022, *About school zones and catchment areas*, [Online], [Cited February 26th, 2024], Available from www.education.sa.gov.au/parents-and-families/enrol-school-or-preschool/zones-catchments-and-capacity-management/about-school-zones-and-catchment-areas.

[36] Department of Basic Education, 2023, *Grade promotion, repetition and dropping out 2018 to 2021*, [Online], [Cited February 26th, 2024], Available from www.education.gov.za.

[37] Department of Basic Education, South Africa, *2023 national senior certificate (nsc) technical report*, accessed January 2024.

[38] Duff A, 2004, *Understanding academic performance and progression of first-year accounting and business economics undergraduates: the role of approaches to learning and prior academic achievement*, Accounting Education, **13(4)**, pp. 409–430.

[39] Education Library, 2021, *Maslow's hierarchy of needs in education*, [Online], [Cited July 8th, 2025], Available from https://educationlibrary.org/maslows-hierarchy-of-needs-in-education/.

[40] Géron A, 2019, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd Edition, O'Reilly Media.

[41] Greenacre M, Groenen P, Hastie T, d'Enza A, Markos A and Tuzhilina E, 2022, *Principal component analysis*, Nature Reviews Methods Primers, **2(1)**, p. 100.

[42] Hoffmann F, Bertram T, Mikut R, Reischl M and Nelles O, 2019, *Benchmarking in classification and regression*, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, **9(5)**, p. e1318.

[43] Hu L and Bentler P, 1999, *Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives*, Structural Equation Modeling, **6**, pp. 1–55.

[44] Huang S, Cai N, Pacheco P, Narrandes S, Wang Y and Xu W, 2018, *Applications of support vector machine (svm) learning in cancer genomics*, Cancer genomics & proteomics, **15(1)**, pp. 41–51.

[45] Imandoust S and Bolandraftar M, 2013, *Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background*, International journal of engineering research and applications, **3(5)**, pp. 605–610.

[46] International Association for the Evaluation of Educational Achievement, 2023, *Pirls: Progress in international reading literacy study*, [Online], [Cited February 6th, 2024], Available from `www.timssandpirls.bc.edu/pirls-landing.html`.

[47] International Association for the Evaluation of Educational Achievement, 2023, *Timss: Trends in international mathematics and science study*, [Online], [Cited February 6th, 2024], Available from `www.timssandpirls.bc.edu/timss-landing.html`.

[48] Jeong D, Aggarwal S, Robinson J, Kumar N, Spearot A and Park D, 2023, *Exhaustive or exhausting? evidence on respondent fatigue in long surveys*, Journal of Development Economics, **161**, p. 102992.

[49] Jiang T, Gradus J and Rosellini A, 2020, *Supervised machine learning: a brief primer*, Behavior therapy, **51(5)**, pp. 675–687.

[50] Joreskog K, 1973, *General methods for estimating a linear structure equation system*, pp. 85–112 in Goldberger AS and Duncan OD (Eds), *Structural Equation Models in the Social Sciences*, pp. 85–112. Seminar Press, New York.

[51] Kaiser H, 1960, *The application of electronic computers to factor analysis*, Educational and Psychological Measurement, **20**, pp. 141–151.

[52] Kaiser H, 1974, *An index of factorial simplicity*, Psychometrika, **39**, pp. 31–6.

[53] Keith D, Yadama A, O'Neill E and Chung S, 2024, *Anticipating the side effects of educational reform using system dynamics modeling*, Review of Research in Education, **48(1)**, pp. 1–27, Available from `https://doi.org/10.3102/0091732X241260012`.

[54] Kline P, 2014, *An easy guide to factor analysis*, Routledge.

[55] Kline R, 2005, *Principles and practice of structural equation modeling*, 2$^{\text{nd}}$ Edition, Guilford Press, New York (NY).

[56] Liew Y, Tan A, EHY, Lim C, Majeed AA, Zhu Y, Chen W, Chen S and Lo J, 2024, *Systems thinking on artificial intelligence integration into higher education: Causal loops*, chapter 6 in López-Ruiz R (Ed), *Complex Systems with Artificial Intelligence*, chapter 6. IntechOpen, Rijeka, Available from `https://doi.org/10.5772/intechopen.1008246`.

[57] Loh W, 2011, *Classification and regression trees*, Wiley interdisciplinary reviews: data mining and knowledge discovery, **1(1)**, pp. 14–23.

[58] Lubisi R and Murphy R, 2002, *Assessment in south african schools*, Assessment in Education: Principles, Policy & Practice, **9(2)**, pp. 255–268.

[59] Makhtar M, Nawang H and Wan Shamsuddin S, 2017, *Analysis on students' performance using naïve bayes classifier*, Journal of Theoretical & Applied Information Technology, **95(16)**.

[60] Maluleke R, 2023, *Quarterly labour force survey (qlfs) q2:2023*, [Online], [Cited February 26th, 2024], Available from `www.statssa.gov.za`.

[61] Maslow A and Lewis K, 1987, *Maslow's hierarchy of needs*, Salenger Incorporated, **14(17)**, pp. 987–990.

[62] Moloi M and Chetty M, *A study of the conditions of schooling and the quality of education the sacmeq iii project in south africa: A study of the conditions of schooling and the quality of education*.

[63] Mullis I, 2007, *PIRLS 2006 International Report: IEA's Progress in International Reading Literacy Study in primary school in 40 countries*, TIMSS & PIRLS International Study Center, Boston College, Newton (MA).

[64] Nasteski V, 2017, *An overview of the supervised machine learning methods*, Horizons. b, **4(51-62)**, p. 56.

[65] Noorbakhsh-Sabet N, Zand R, Zhang Y and Abedi V, 2019, *Artificial intelligence transforms the future of health care*, The American journal of medicine, **132(7)**, pp. 795–801.

[66] Oliver R and Bearden W, 1985, *Crossover effects in the theory of reasoned action: A moderating influence attempt*, Journal of consumer research, **12(3)**, pp. 324–340.

[67] O'Rourke N and Hatcher L, 2013, *A Step-by-Step Approach to Using SAS for Factor Analysis and Structural Equation Modeling*, 2nd Edition, SAS Institute Inc., Cary, NC, Available from `http://support.sas.com/bookstore`, accessed: 2025-07-08.

[68] Peng C, So T, Stage F and St John E, 2002, *The use and interpretation of logistic regression in higher education journals: 1988–1999*, Research in higher education, **43**, pp. 259–293.

[69] Quené H and Van den Bergh H, 2004, *On multi-level modeling of data from repeated measures designs: A tutorial*, Speech communication, **43(1-2)**, pp. 103–121.

[70] Ramasu T and Kanakana-Katumba G, 2025, *Evaluating the dynamics of fee-free higher education in south africa: a causal loop diagram approach*, F1000Research, **13**, p. 780, Available from `https://doi.org/10.12688/f1000research.152478.3`.

[71] Rasmussen University, 2024, *Understanding maslow's hierarchy of needs in education*, [Online], [Cited July 8th, 2025], Available from `https://www.rasmussen.edu/degrees/education/blog/understanding-maslows-hierarchy-of-needs-in-education/`.

[72] Reddy V, Winnaar L, Juan A, Arends F, Harvey J, Hannan S, Namome C, Sekhejane P and Zulu N, 2019, *Timms 219 highlights of south african grade 9 results in mathematics and science*, Available from `www.education.gov.za`.

[73] Research on Socio-Economic Policy, 2023, *Research on socio-economic policy*, [Online], [Cited February 8th, 2024], Available from `www.resep.sun.ac.za/`.

[74] Saidani O, Menzli L, Ksibi A, Alturki N and Alluhaidan A, 2022, *Predicting student employability through the internship context using gradient boosting models*, Ieee Access, **10**, pp. 46472–46489.

[75] Salazar F, Toledo M, Oñate E and Morán R, 2015, *An empirical comparison of machine learning techniques for dam behaviour modelling*, Structural Safety, **56**, pp. 9–17.

[76] Samuel A, 1959, *Some studies in machine learning using the game of checkers*, IBM Journal of research and development, **3(3)**, pp. 210–229.

[77] SAS Institute, 2024, *Sas*, [Online], [Cited June 21st, 2024], Available from www.sas.com/en_za/home.html.

[78] Schreiber J, 2021, *Issues and recommendations for exploratory factor analysis and principal component analysis*, Research in Social and Administrative Pharmacy, **17(5)**, pp. 1004–1011.

[79] Sharma P and Kim K, 2013, *A comparison of pls and ml bootstrapping techniques in sem: A monte carlo study*, Proceedings of the New perspectives in partial least squares and related methods, Houston, pp. 201–208.

[80] Shaukat K, Luo S, Varadharajan V, Hameed I, Chen S, Liu D and Li J, 2020, *Performance comparison and current challenges of using machine learning techniques in cybersecurity*, Energies, **13(10)**, p. 2509.

[81] Shimp T and Kavas A, 1984, *The theory of reasoned action applied to coupon usage*, Journal of consumer research, **11(3)**, pp. 795–809.

[82] Sinha P, Calfee C and Delucchi K, 2021, *Practitioner's guide to latent class analysis: methodological considerations and common pitfalls*, Critical care medicine, **49(1)**, pp. e63–e79.

[83] Slamang S, 2020, *A systems perspective on public high school management within the western cape*, Master of Commerce Thesis, Stellenbosch University, Stellenbosch.

[84] SMF News, 2024, *Athletes raise r2.3 million for low-income learners*, [Online], [Cited April 2nd, 2025], Available from https://www.smfnews.org/athletes-raise-r2-3-million-for-low-income-learners/.

[85] South African Government, 1994, *President nelson mandela: 1994 presidential inauguration*, [Online], [Cited February 2nd, 2024], Available from www.gov.za/news/speeches/president-nelson-mandela-1994-presidential-inauguration-10-may-1994.

[86] South African Government, 2021, *Curriculum assessment policy statements (caps)*, [Online], [Cited February 2nd, 2024], Available from www.education.gov.za/Curriculum/CurriculumAssessmentPolicyStatements(CAPS).aspx.

[87] South African Government, 2021, *Nsc examinations*, [Online], [Cited February 2nd, 2024], Available from www.education.gov.za/Curriculum/NationalSeniorCertificate(NSC)Examinations.aspx.

[88] South African Government, 2023, *Department of basic education*, [Online], [Cited January 31st, 2024], Available from www.education.gov.za/AboutUs/AboutDBE.aspx#:~:text=Our%20mission%20is%20to%20provide,system%20for%20the%2021st%20century.&text=People%3A%20Upholding%20the%20Constitution%2C%20being,the%20people%20of%20South%20Africa.

[89] South African Government, 2023, *Department of higher education and training (dhet)*, [Online], [Cited January 31st, 2024], Available from www.nationalgovernment.co.za/units/view/17/department-of-higher-education-and-training-dhet#:~:text=The%20Department's%20mission%20is%20to,development%20goals%20of%20the%20country.

[90] South African Government, 2023, *Statistics south africa on quarterly labour force survey quarter three 2023*, [Online], [Cited February 2nd, 2024], Available from www.gov.za/news/media-statements/statistics-south-africa-quarterly-labour-force-survey-quarter-three-2023-14#:~:text=This%20resulted%20in%20a%20net,the%20third%20quarter%20of%202023.

[91] South African Schools Act of 1966, south Africa, [Statute], Government Gazette, Vol. 377 No. 17579, Pretoria.

[92] Spaull N, 2013, *Poverty & privilege: Primary school inequality in south africa*, International Journal of Educational Development, **33(5)**, pp. 436–447.

[93] Spearman C, 1904, *General intelligence objectively determined and measured*, American Journal of Psychology, **15**, pp. 201–293.

[94] Stellenbosch University, 2025, (Unpublished) , Technical report, Stellenbosch University, Available from https://www.sun.ac.za/english/maties/Documents/Admission%20booklet%202025%20%28English%29.pdf.

[95] Systems Thinking for Education Policy, 2023, *Systems thinking for education policy*, [Online], [Cited February 8th, 2024], Available from www.step.org.za/.

[96] Tavakol M and Wetzel A, 2020, *Factor analysis: a means for theory and instrument development in support of construct validity*, International journal of medical education, **11**, p. 245.

[97] van den Heever M, Becker E, Venter L and Bekker J, 2024, *Using machine learning and agent-based simulation to predict learner progress for the south african high school education system*, South African Journal of Industrial Engineering, **35(3)**, pp. 15–27.

[98] van der Berg S, Hoadley U, Galant J, van Wyk C and Böhmer B, 2022, *Learning losses from covid-19 in the western cape: Evidence from systemic tests*, Research on Socio Economic Policy (Resep), Stellenbosch University February.

[99] Venter L, *A systems perspective of basic education in south africa.*

[100] Venter L, 2024, *Statistical analysis for variate relationships in the south african ecd system.*

[101] Watkins W, 2018, *Exploratory factor analysis: A guide to best practice*, Journal of black psychology, **44(3)**, pp. 219–246.

[102] Western Cape Government, 2019, *Schools of skills*, [Online], [Cited February 26th, 2024], Available from www.westerncape.gov.za/general-publication/school-skills.

[103] Wills G and Qvist J, 2023, (Unpublished) , Covid-generation research report, Research on Socio-Economic Policy (RESEP), Stellenbosch University, prepared with support from Allan and Gill Gray Philanthropies.

APPENDIX A

Supplementary material

A.1 Variable aggregation

This section outlines how variables from the GHS were recoded into analytical indicators. The variables are grouped as personal, household, and school-specific characteristics.

A.1.1 Personal characteristics for GHS

Table A.1: *Variable aggregation for selected personal characteristics from the GHS.*

Variable name	GHS categories	Indicator variable
EDU_MODE_TR	Walking	Walk to school
EDU_FOOD	Yes	School feeding
EDU_ATTEND	Yes	Retention
DSB_WAL	No difficulty	Good walking
DSB_REM	No difficulty	Good memory
DSB_SEL	No difficulty	Good selfcare
DSB_COM	No difficulty	Good communication
HHC_FATH_PARTHH	Is respondent's biological father part of household	Paternal participation
HHC_MOTH_PARTHH	Is respondent's biological father part of household	Maternal participation
HHC_RELATIONSHIP	Son/daughter/stepchild/adopted child of household head	Co-residence with parents

A.1.2 Household characteristics for GHS

Table A.2: *Variable aggregation for selected household characteristics from the GHS.*

Variable name	GHS categories	Indicator variable
SOC_GRANT	Yes	Grant receiver
SOC_GRANT_TYPE	Child support grant	Child support grant
FIN_INC	Remittances (money received from people living elsewhere), pensions, grants (include old age grant here)	Unearned income
ENG_ACCESS	Yes	Energy access
ENG_MAINELECT	In-house conventional meter, in-house prepaid meter	Main electricity meter
ENG_PAY	Yes	Paid electricity

TABLE A.2: *Variable aggregation for selected household characteristics from the GHS (continued).*

Variable name	**GHS categories**	**Indicator variable**
`SWR_RUB`	Removed by local authority/private company at least once a week, removed by local authority/private company less often than once a week, removed by community members, contracted by the municipality, at least once per week, removed by community members, contracted by the municipality, less often than once per week, removed by community members at least once a week, removed by community members less often than once a week	`Managed waste collection`
`WAT_DRINKWAT`	Piped (tap) water in dwelling/house, piped (tap) water in yard	`In-house water`
`SAN_TOIL`	Flush toilet connected to a public sewerage system, flush toilet connected to a septic or conservancy tank, pour bucket-flush toilet connected to a septic tank (or seepage pit), bucket toilet	`General toilet`
`SWR_ENV_LIT`	No	`No littering`
`SWR_ENV_WAT`	No	`No water pollution`
`SWR_ENV_AIR`	No	`No air pollution`
`SWR_ENV_NOI`	No	`No noise pollution`
`FSD_RANOUT`	No	`Food available`
`FSD_ATELESS`	No	`Food sufficient`
`FSD_FEWFOODS`	No	`Food variety`
`FSD_HUNGRY`	No	`Household not hungry`

A.1.3 Focus school characteristics

TABLE A.3: *Variable aggregation for the focus school*

Indicator	**Question**
Father's emotional support	Does your father emotionally engage with you such as having regular conversation with you or checking in to see how you are doing?
Father's physical support	Is your father physically present at important occasions or when you need his help?
Father's financial support	Does your father contribute financially to your life such as paying for food, accommodation, or school fees?
Co-residence with parents	Do you live with both a mother figure and a father figure?
School satisfaction	Do you feel you get a good quality of teaching at school?
Future enthusiasm	Does your school make you excited about your future?
School motivation	Are you motivated to attend school?
Successful life preparation	Do you feel that your school is setting you up for success in life?
Teacher availability	Does your school have enough teachers for all learners?
Learning materials	Do you have enough learning materials (such as desks, white boards, chairs) and books at school?
School infrastructure	Is the school building in good condition?
School values promotion	Does your school promote good values?

A.2 Determining the number of factors

This section describes how the number of latent factors was determined for each group of schools using scree plots based on eigenvalues from exploratory factor analysis.

A.2.1 Scree plots for SQ13 schools

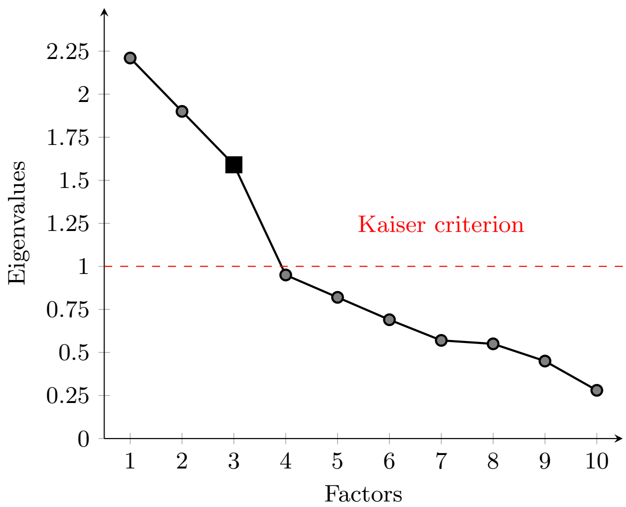

Figure A.1: *Scree plot showing eigenvalues for determining the appropriate number of factors to retain during EFA of person factors for SQ13 schools. The plot illustrates both the Kaiser criterion (eigenvalue > 1) and the elbow method, both suggesting 3 factors should be retained.*

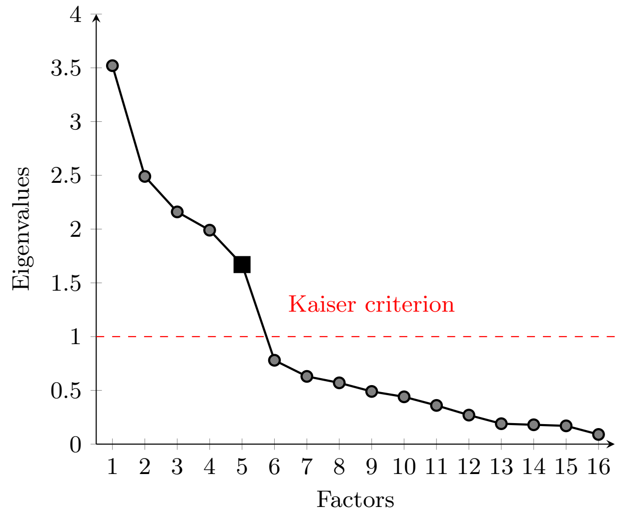

Figure A.2: *Scree plot showing eigenvalues for determining the appropriate number of factors to retain during EFA of household factors for SQ13 schools. The plot illustrates both the Kaiser criterion (eigenvalue > 1) and the elbow method, both suggesting 5 factors should be retained.*

A.2.2 Scree plots for SQ45 schools

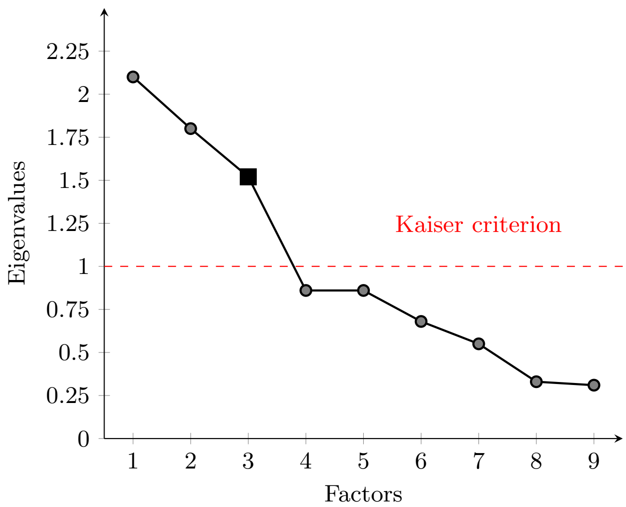

FIGURE A.3: *Scree plot showing eigenvalues for determining the appropriate number of factors to retain during EFA of person factors for SQ45 schools. The plot illustrates both the Kaiser criterion (eigenvalue > 1) and the elbow method, both suggesting 3 factors should be retained.*

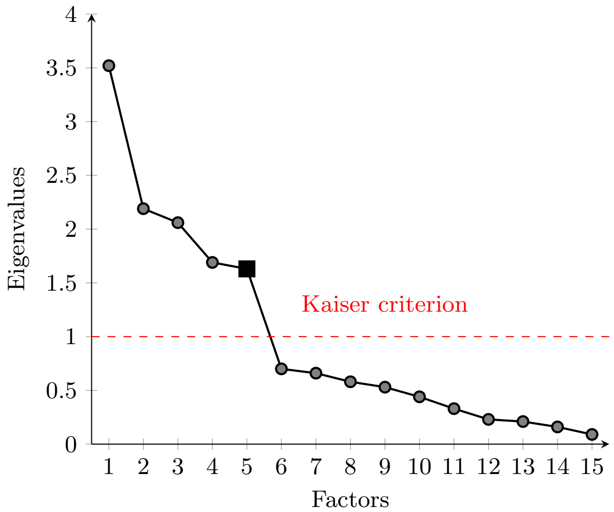

FIGURE A.4: *Scree plot showing eigenvalues for determining the appropriate number of factors to retain during EFA of household factors for SQ45 schools. The plot illustrates both the Kaiser criterion (eigenvalue > 1) and the elbow method, both suggesting 5 factors should be retained.*

A.2.3 Scree plots for the focus school

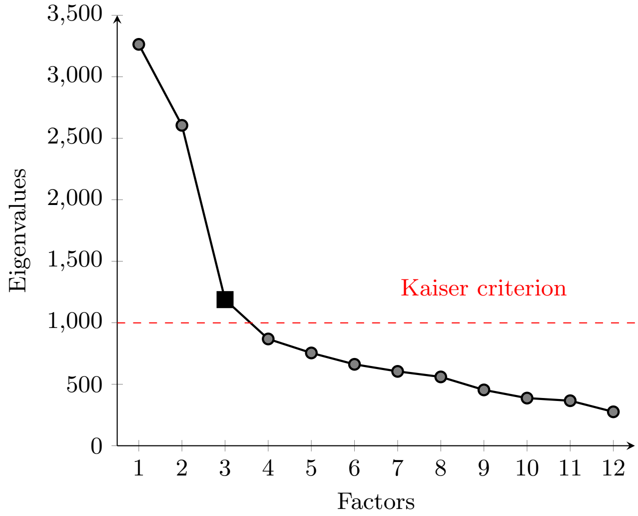

FIGURE A.5: *Scree plot showing eigenvalues for determining the appropriate number of factors to retain during EFA of learner factors for the focus school. The plot illustrates both the Kaiser criterion (eigenvalue > average eigenvalue of 999) and the elbow method, both suggesting 3 factors should be retained.*

A.3 Granular MRA and SEM analysis

This section presents the results of MRA and SEM for selected indicators, analysed separately by school group. The aim is to show how household and learner characteristics relate to one another within each setting.

A.3.1 LQ13 learners in SQ13 schools

	Grant receiver	Child support grant	Unearned income	In-house water	General toilet	Managed waste collection	Walk to school	Retention	School feeding	Good walking	Good memory	Good hygiene	Good communication	Co-residence with parents	Maternal survival status	Paternal participation	Maternal participation
Grant receiver	1.00	0.88**	0.12**	-0.01	-0.00	-0.02	0.01	0.01	-0.01	0.05	-0.11	0.12	0.00	0.02	-0.03*	-0.01	0.00
Child support grant	0.88**	1.00	0.03	0.01	0.00	0.03*	-0.02	0.09**	0.03	-0.01	0.18**	-0.22	-0.03	-0.02	0.06**	0.01	0.01
Unearned income	0.12**	0.03	1.00	-0.06**	-0.02	-0.01	0.03*	-0.17**	0.00	0.06	-0.07	-0.01	0.05	-0.06**	-0.01	-0.05**	0.00
In-house water	-0.01	0.01	-0.06**	1.00	0.45**	0.16**	-0.06**	0.01	-0.04	-0.18	0.12	0.09	-0.01	-0.06**	0.00	0.08**	0.03
General toilet	-0.00	0.00	-0.02	0.45**	1.00	0.65**	0.05**	-0.03	-0.02	0.06	-0.05	-0.04	0.03	0.04**	0.00	-0.04**	-0.01
Managed waste collection	-0.02	0.03*	-0.01	0.16**	0.65**	1.00	-0.01	0.05	-0.04	0.06	-0.10	0.10	-0.04	-0.02	-0.01	0.02	0.00
Walk to school	0.01	-0.02	0.03*	-0.06**	0.05**	-0.01	1.00	0.69**	0.13**	-0.25	0.23*	-0.15	-0.10	-0.01	-0.03	-0.02	0.01
Retention	0.01	0.09**	-0.17**	0.01	-0.03	0.05	0.69**	1.00	0.35**	-0.14*	0.01	0.18*	0.05	-0.00	0.01	0.01	-0.01
School feeding	-0.01	0.03	0.00	-0.04	-0.02	-0.04	0.13**	0.35**	1.00	0.10	-0.06	0.21	-0.03	0.05*	-0.01	-0.02	0.00
Good walking	0.05	-0.01	0.06	-0.18	0.06	0.06	-0.25	-0.14*	0.10	1.00	0.24*	0.90**	0.19	-0.22	-0.11	-0.24	0.00
Good memory	-0.11	0.18**	-0.07	0.12	-0.05	-0.10	0.23*	0.01	-0.06	0.24*	1.00	0.64**	0.91**	0.05	0.11	0.01	-0.01*
Good hygiene	0.12	-0.22	-0.01	0.09	-0.04	0.10	-0.15	0.18*	0.21	0.90**	0.64**	1.00	0.71**	0.17	-0.10	0.00	0.00
Good communicat	0.00	-0.03	0.05	-0.01	0.03	-0.04	-0.10	0.05	-0.03	0.19	0.91**	0.71**	1.00	-0.04	-0.08	0.01	0.01
Co-residence with parents	0.02	-0.02	-0.06**	-0.06**	0.04**	-0.02	-0.01	-0.00	0.05*	-0.22	0.05	0.17	-0.04	1.00	-0.03	0.35**	0.59**
Maternal survival status	-0.03*	0.06**	-0.01	0.00	0.00	-0.01	-0.03	0.01	-0.01	-0.11	0.11	-0.10	-0.08	-0.03	1.00	-0.00	0.37**
Paternal participatio	-0.01	0.01	-0.05**	0.08**	-0.04**	0.02	-0.02	0.01	-0.02	-0.24	0.01	0.00	0.01	0.35**	-0.00	1.00	-0.06*
Maternal participatio	0.00	0.03	-0.01	0.03	-0.02	0.01	0.01	-0.03	0.00	0.25	-0.17*	-0.04	0.11	0.56**	-0.46**	-0.05	1.00

TABLE A.4: *Regression coefficient matrix for the seventeen indicators based on multivariate regression analysis results for LQ13 learners in SQ13 schools. Rows represent predictors, columns represent outcomes. Cell shading indicates the magnitude of the correlation coefficient. ** indicates $p < 0.001$, * indicates $p < 0.05$.*

Predictor	Outcome	β	SE	t-value
Grant receiver	Child support grant	0.89	0.00	233.80
Retention	School feeding	0.60	0.01	49.65
Co-residence with parents	Maternal participation	0.57	0.01	49.45
Good memory	Good communication	0.30	0.02	17.73
General toilet	Managed waste collection	0.81	0.01	125.60
In-house water	General toilet	0.64	0.01	58.18
Co-residence with parents	Paternal participation	0.32	0.01	23.92
Good walking	Good hygiene	0.40	0.02	25.28
Retention	Child support grant	0.11	0.02	7.34
Child support grant	Maternal survival status	0.12	0.02	6.63
Unearned income	Grant receiver	0.31	0.02	18.68
School feeding	Walk to school	0.27	0.02	15.61
Retention	Unearned income	-0.08	0.02	-5.15
Unearned income	Co-residence with parents	-0.13	0.02	-6.55
Unearned income	Paternal participation	-0.08	0.02	-4.07
Good memory	Retention	0.04	0.02	2.15
School feeding	Child support grant	0.13	0.02	7.50
Unearned income	In-house water	-0.13	0.02	-7.48
Grant receiver	General toilet	-0.03	0.01	-3.92
In-house water	Paternal participation	0.07	0.01	4.91
Managed waste collection	Paternal participation	0.08	0.02	4.33

TABLE A.5: *Standardised effects of predictors on outcomes including the estimated path coefficients (β), standard errors (SE), and t-values for each relationship of the seventeen indicators for LQ13 learners in SQ13 schools.*

A.3.2 LQ13 learners in SQ45 schools

	Grant receiver	Childcare grant	Unearned income	In-house water	General toilet	Managed waste collection	Retention	School feeding	Walking to school	Co-residence with parents	Paternal participation	Maternal participation
Grant receiver	1.00	0.92**	0.07**	-0.00	-0.02	0.01	-0.20**	0.01	-0.07	-0.01	-0.00	-0.02
Childcare grant	0.92**	1.00	0.02	0.00	0.04*	-0.02	0.22**	0.04**	0.11	0.01	-0.00	0.03*
Unearned income	0.24**	0.09	1.00	-0.09**	-0.09**	0.00	-0.32**	0.10**	0.07*	-0.17**	-0.14**	0.02
In-house water	-0.01	0.00	-0.09**	1.00	0.46**	0.24**	0.18**	-0.06*	-0.07*	-0.08**	0.05*	0.06*
General toilet	-0.05	0.09*	-0.05**	0.27**	1.00	0.64**	-0.03	-0.03*	0.02	0.01	0.01	-0.00
Managed waste collection	0.02	-0.04	0.00	0.15**	0.60**	1.00	0.04	-0.06**	0.02	-0.01	0.01	0.03
Retention	-0.21**	0.22**	-0.09**	0.06**	-0.02	0.02	1.00	0.09**	0.25**	-0.03*	-0.00	0.02
School feeding	0.06	0.19**	0.11**	-0.07**	-0.07*	-0.12**	0.37**	1.00	0.38**	-0.00	0.01	0.04
Walking to school	-0.07	0.11	0.07*	-0.07*	0.02	0.02	0.25**	0.38**	1.00	0.00	-0.05*	-0.03
Co-residence with parents	-0.03	0.03	-0.14**	-0.08**	0.02	-0.01	-0.09*	-0.00	0.00	1.00	0.31**	0.62**
Paternal participation	-0.00	-0.01	-0.11**	0.06*	0.02	0.01	-0.00	0.01	-0.05*	0.46**	1.00	-0.02
Maternal participation	-0.06	0.12*	0.02	0.06*	-0.01	0.05	0.05	0.04	-0.03	0.62**	-0.02	1.00

Table A.6: *Regression coefficient matrix for the twelve indicators based on multivariate regression analysis results for LQ13 learners in SQ45 schools. Rows represent predictors, columns represent outcomes. Cell shading indicates the magnitude of the correlation coefficient. ** indicates* $p < 0.001$, ** indicates* $p < 0.05$.

Predictor	Outcome	β	SE	t-value
Grant receiver	Child care grant	0.90	0.00	186.80
Maternal participation	Co-residence with parents	0.53	0.02	33.57
Paternal participation	Co-residence with parents	0.34	0.02	19.60
Grant receiver	Unearned income	0.38	0.02	19.42
Managed waste collection	General toilet	1.01	0.05	18.65
Walk to school	School feeding	0.36	0.02	18.28
Paternal participation	Maternal participation	0.32	0.02	14.69
General toilet	In-house water	0.58	0.05	12.14
Child care grant	School feeding	0.24	0.02	11.25
Child care grant	Retention	0.49	0.05	10.05
Grant receiver	Retention	-0.44	0.05	-8.91
School feeding	Retention	0.21	0.03	7.86
Co-residence with parents	Unearned income	-0.19	0.02	-7.83
Child care grant	Walk to school	0.18	0.02	7.47
Managed waste collection	In-house water	0.34	0.05	7.15
Unearned income	Retention	-0.17	0.02	-7.10
Paternal participation	Unearned income	-0.16	0.02	-6.56
Paternal participation	Grant receiver	-0.15	0.02	-6.15
Paternal participation	Managed waste collection	0.13	0.02	5.87
In-house water	Retention	0.13	0.02	5.86
Walk to school	Retention	0.12	0.02	4.95
Maternal participation	Managed waste collection	0.10	0.02	4.72
Unearned income	School feeding	0.09	0.02	4.27
Maternal participation	Child care grant	0.04	0.01	3.99
In-house water	Walk to school	-0.09	0.02	-3.89
Unearned income	Walk to school	0.10	0.03	3.86

TABLE A.7: *Standardised effects of predictors on outcomes including the estimated path coefficients (β), standard errors (SE), and t-values for each relationship of the twelve indicators for LQ13 learners in SQ45 schools.*